View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Illinois Digital Environment for Access to Learning and Scholarship Repository

REGULATION OF FLOURY2-mRFP EXPRESSION IN RESPONSE TO LONG TERM SELECTION FOR GRAIN PROTEIN CONCENTRATION IN MAIZE

BY

CHRISTINE JEANETTE LUCAS

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Master's Committee:

Associate Professor Stephen P. Moose, Chair
Professor Steven C. Huber
Professor Raymond E. Zielinski
Professor A. Lane Rayburn

ABSTRACT

The Illinois Long-term Selection Experiment is the longest running continuous genetics experiment in higher plants in the world. Seven hundred and twenty-one cycles of divergent selection for grain protein and oil concentration have produced populations with the known phenotypic extremes for this trait, and also illustrate the nature of responses to phenotypic selection. This report provides updates to the experiment, including the introduction of new chemical analytical procedures, assessment of methods for estimating genetic gain, and the recent initiation of reverse selection experiments from cycle 103 of Illinois High Protein (IHP). Prior genetic mapping studies suggest that the response to selection in this experiment is dependent upon the cumulative action of many genes with small phenotypic effects. An alternative theory explored here is that the response depends on quantitative expression variation of a few major regulators. Analysis of crosses between IHP and Illinois Reverse High Protein (IRHP) provides evidence that the response to selection in IRHP is due to one or a few loci with main effects, which is associated with biased frequencies for variant alleles of the *ASPARAGINE SYNTHETASE3* gene.

Long-term divergent selection for grain protein concentration in the Illinois Protein Strains has dramatically altered the accumulation of the 19-kD and 22-kD α-zeins. Known regulators of α-zein protein accumulation are *Opaque2* (*O2*), the *Prolamin-box Binding Factor* (*PBF*), and factors influencing the folding of zeins into endosperm protein bodies. RNA analysis and measuring protein abundance are two effective approaches for studying the regulation of zein expression, but they are also expensive, destructive and laborious. Furthermore, study of individual zein genes is complicated by their high copy number and sequence similarity. An alternative inexpensive and nondestructive approach to investigate the regulation of zein expression is the use of transgenic *Floury2-mRFP* reporter lines (Dave Jackson's lab at Cold Spring Harbor), where the expression of the readily visible monomeric red fluorescent protein (mRFP) is controlled by the genomic sequences encoding the

Floury2 **α-zein. The *FL2-mRFP* transgene has been introgressed into inbred lines derived from the four Illinois Protein Strains (IPS), as well as the reference inbred B73. We found that *FL2-mRFP* expression not only correlates with grain protein concentration, but also follows known patterns of zein accumulation throughout development. At all developmental stages, RFP expression was strongest in Illinois High Protein (IHP), the lowest in Illinois Low Protein (ILP) and intermediate in Illinois Reverse High Protein (IRHP), Illinois Reverse Low Protein (IRLP), and B73. By crossing *FL2-mRFP* to an *o2* mutant introgressed into IHP, we show that its expression is strongly activated by *O2*, illustrating that the *FL2-mRFP* transgene is regulated in the same manner as endogenous α-zein genes. Future experiments will use the *FL2-mRFP* transgene as a tool for identifying regulators influencing protein concentration in ongoing genetic mapping studies.**

TABLE OF CONTENTS

CHAPTER 1

THE ILLINOIS LONG-TERM SELECTION EXPERIMENT

Abstract

The Illinois Long-term Selection Experiment is the longest running continuous genetics experiment in higher plants in the world. Seven hundred and twenty-one cycles of divergent selection for grain protein and oil concentration have produced populations with the known phenotypic extremes for this trait, and also illustrate the nature of responses to phenotypic selection. This report provides updates to the experiment, including the introduction of new chemical analytical procedures, assessment of methods for estimating genetic gain, and the recent initiation of reverse selection experiments from cycle 103 of Illinois High Protein (IHP). Prior genetic mapping studies suggest that the response to selection in this experiment is dependent upon the cumulative action of many genes with small phenotypic effects. An alternative theory explored here is that the response depends on quantitative expression variation of a few major regulators. Analysis of crosses between IHP and Illinois Reverse High Protein (IRHP) provides evidence that the response to selection in IRHP is due to one or a few loci with major effects, which is associated with biased frequencies for variant alleles of the *ASPARAGINE SYNTHETASE3* gene.

INTRODUCTION

Illinois Protein and Oil Strains

Initiated in 1896 at the University of Illinois, Cyril G. Hopkins analyzed grain protein and oil concentrations from 163 ears of the open-pollinated maize variety, "Burr's White". While the initial objective was to determine if selection could alter the chemical composition of the kernel, it was later modified to determine the limits of selection for protein and oil grain concentrations. The 24 ears with the highest protein concentrations were selected as the parents for Illinois High Protein (IHP), the 24 with the highest oil concentrations for Illinois High Oil (IHO), and the lowest 12 ears for these traits became parents of Illinois Low Protein (ILP) and Illinois Low Oil (ILO). Recurrent selection has been conducted on these four populations for 110 cycles, making the experiment the longest running continuous genetics experiment in higher plants. The resulting four populations span the known extremes for grain protein and oil compositions. Compared to an average of 8-12% grain protein, IHP kernels contain over 32% protein and ILP kernels only 4% protein (**Figure 1a**). Similarly, while typical maize kernels contain an average of 4-6% oil, IHO kernels contain more than 20% oil and ILO kernels 1% oil (**Figure 1b**).

In order to determine the extent of genetic variability and phenotypic response remaining after 48 generations of forward selection, Dr. Earl Leng reversed the direction of selection in IHP, ILP, IHO and ILO populations. Although the method of selection was the same, it was conducted in the opposite direction as the original forward strains. Low protein ears were selected from IHP to create Illinois Reverse High Protein (IRHP) and high protein ears were selected from ILP to create Illinois Reverse Low Protein (IRLP) (**Figure 1a**). Illinois Reverse High Oil (IRHO) and Illinois Reverse Low Oil (IRLO) were created in the same way from IHO and ILO, respectively (**Figure 1b**). Illinois Switchback High Oil (ISHO) was initiated from cycle 55 of IRLO with the same overall goal of determining genetic variation. To determine the extent of variability remaining in ILP when it was discontinued, the Illinois Reverse Low

Protein 2 (IRLP2) selection was initiated from cycle 90 of ILP. Two new reverse selections are introduced here, Illinois Reverse High Protein 2 and 3 (IRHP2 and IRHP3). IRHP2 and IRHP3 were each initiated from cycle 103 of IHP by selecting for low protein individuals. They were created for three reasons: 1) to estimate the extent of genetic variability remaining in IHP after more than 100 cycles of forward selection, 2) to assess the impact of soil N fertility on phenotypic means and selection response, and 3) to initiate new replicated populations where DNA of the founding individuals is preserved for future analysis of genetic responses to selection.

Chemical Analytical Procedures

Previous chemical analytical procedures for measuring protein and oil are summarized in Dudley and Lambert (2004). The protein analytical procedure was changed in 2006 from a combustion analyzer (CE Elantech Inc. NA2000 N-Protein, Lakewood, NJ) using the method of Dumas (Kirsten, 1983) to a Near Infrared Reflectance machine (NIR, DICKEY-john Instalab 600) using the method of Dudley and Lambert (1992). One disadvantage of N combustion is that it only measures protein concentrations and oil had to be measured separately utilizing Nuclear Magnetic Resonance (NMR). The use of NIR provides a potential advantage by measuring not only protein and oil concentrations, but also starch, moisture, and possibly other components where calibrations have been established, thereby not only eliminating the need for multiple methods, but generating more data as well. Here we analyze the suitability of using NIR as method for measuring grain composition in the Illinois Long Term Selection Experiment.

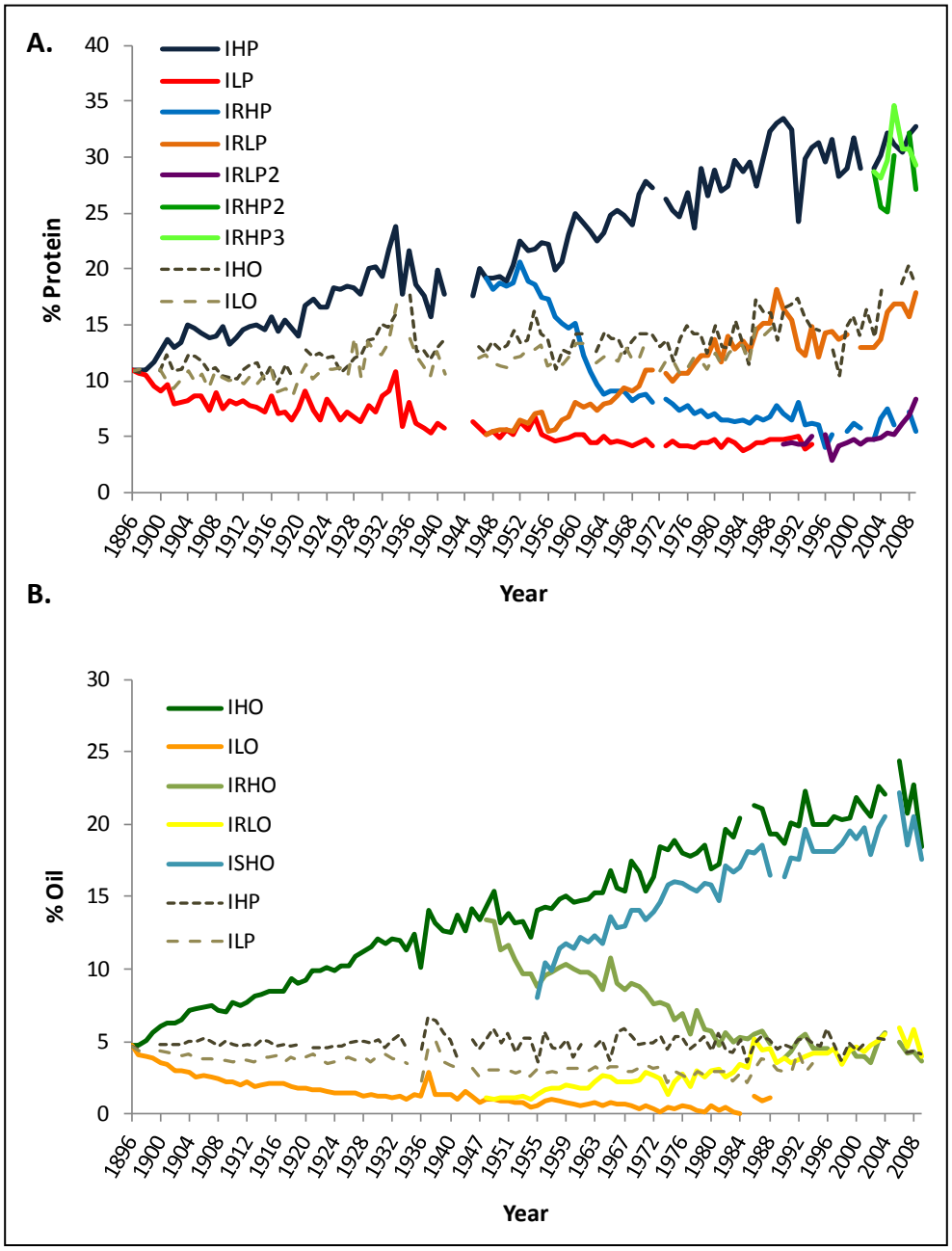


Figure 1. Selection responses in the a.) Illinois Protein Strains and b.) the Illinois Oil Strains. 109 generations of recurrent selection have created Illinois High Protein (IHP), Illinois Low Protein (ILP), Illinois High Oil (IHO) and Illinois Low Oil (ILO) strains. After 48 cycles of forward selection, the direction of selection was reversed in each of these four strains to produce Illinois Reverse High Protein (IRHP), Illinois Reverse Low Protein (IRLP), Illinois Reverse High Oil (IRHO) and Illinois Reverse Low Oil (IRLO). Illinois Reverse Low Protein 2 (IRLP2) was initiated from ILP at cycle 90, and Illinois Reverse High Protein 2 (IRHP2) and Illinois Reverse High Protein 3 (IRHP3) were initiated from IHP at cycle 103. Switchback High Oil (SHO) was initiated from IRLO at generation 55. Protein concentrations are plotted for IHO and ILO in (a), and oil concentrations are plotted for IHP and ILP (b).

Assessment of Methods for Calculating Genetic Gain in the Protein and Oil Strains, and Generation of Novel Methods for Estimating Environmental Variation

One long-term objective of the ILTSE is to determine the limits to selection for grain protein concentration in the Illinois Protein Strains and oil concentration in the Illinois Oil Strains. Limits have clearly been reached for two of the strains; ILP has been discontinued due to poor germination and ILO due to a lack of methods for detecting such low (<1%) oil concentrations. However, there is evidence that a limit has not been reached for other strains. The high degree of progress made in all of the original reverse strains (IRHP, IRLP, IRHO, IRLO) demonstrates that forty-eight cycles of forward selection had not diminished genetic variability of the forward strains (IHP, ILP, IHO, ILO). In order to determine the extent of variability remaining after one-hundred generations of selection, realized heritability and gain per generation was calculated for all of the strains (Dudley and Lambert, 2004). The results of this study showed that IHO and SHO still exhibited significant change per generation and realized heritability. The results were less clear for IHP because, although realized heritability and genetic variance were significant, change per generation was not. Tests for genetic variability within recent cycles of ILP were not possible because it was discontinued at cycle 90, but could be estimated by the progress of IRLP2. However, the results for IRLP2 were inconclusive. Thus, it is unclear as to whether the limits to selection have been definitively reached for all strains. Here, we analyze the strains over the last decade (2000 – 2009) to determine if progress is still being made.

Dudley and Lambert (2004) divided the selection experiment into segments based on changes in selection procedure and environmental conditions that may have affected the rate of progress. Gain per segment was calculated by regressing generation means on cumulative selection differential. However, this method for calculating gain does not take into account the extent of phenotypic variation due to the environment, which becomes more important when the rate of gain begins to plateau, as seen in the later generations of IHP, for example. In these later generations, environmental variation

represents a larger proportion of the observed gain, and it becomes more difficult to determine if changes in protein are due to selection or the environment. In future generations, it will be necessary to more precisely distinguish genetic from environmental factors in order to determine if progress is still being made.

One environmental factor that is expected to have a significant impact on the grain protein concentration phenotype is soil nitrogen level, as it is known that cereal seed protein concentrations increase following nitrogen fertilization. Grain protein concentrations were observed to increase with N supply in hybrids derived from the Illinois Protein Strains (Uribelarrea et al., 2004), particularly for IHP. Nitrogen fertilizer was first applied to plots for the Illinois Selection Experiment beginning in generation 53 for IHP and ILP and generation 59 for IHO and ILO. The use of N fertilizer was proposed to explain the increased rate of gain in IHP in the segment in which fertilizer was applied (segment 4) compared to previous segments (Dudley and Lambert, 2004). Although the gain from selection observed in IHP during segment 4 has been attributed to fertilizer application, the effect of N on protein concentration in the IPS has never been directly tested.

Physiological and Genetic Responses to Long-term Selection for Grain Protein Concentration

Selection for grain protein concentration has not only affected grain composition, as it was quickly observed that selection operated indirectly on other traits (Hopkins, 1903; Smith, 1908). Grain starch concentration and kernel size were shown to be inversely related to grain protein concentration in all of the Illinois Protein Strains. Indirect selection has also influenced grain yield, which has largely been attributed to the strong negative correlation between grain protein and starch concentrations, and the positive correlation of grain starch with grain yield. Whereas IHP contains the highest grain protein concentration, it has the lowest starch concentration and the lowest yield. In contrast, ILP contains the lowest grain protein concentration, but the highest starch concentration and the highest yield. On

average, ILP contains only half of the protein as IHP, but two-fold more grain yield than IHP (Below et al., 2004). Since kernel size is also negatively correlated with grain protein concentration, IHP kernels are much smaller, and although they contain twice as much protein as ILP kernels, their grain protein concentration is much higher than a two-fold increase over ILP (Below et al., 2004). Selection for high protein has also resulted in greater lodging and shorter plant height, and increases in successful germination and tillering (Woodworth et al., 1952). These traits are oppositely affected by selection for low protein, where the poor germination of ILP was one of the reasons it was discontinued.

Whole-plant nitrogen metabolism may also be altered by selection for grain protein concentration. The N assimilated by the plant during vegetative development accumulates as protein in the kernel where it is stored until needed as a source of N by the next generation developing seedling (source). Physiological changes affecting nitrogen metabolism in response to selection for grain protein concentration were initially noted by Hoener and DeTurk (1938) and have been reviewed by Below et al. (2004). The results of these studies demonstrate elevated N uptake, N assimilation by seedling leaves, and N remobilization from source to seed sink tissues of IHP compared to ILP. Uribellarrea et al. (2004 and 2007) also found that when the inbred-derived Illinois Protein Strains were crossed to an elite common tester, the hybrids exhibited a more than two-fold difference in grain protein that is attributed to differences in nitrogen use efficiency. IHP was also more efficient at remobilizing N than IRLP. One underlying cause of these physiological differences is changes in the activities of the enzymes involved in these pathways. For example, IHP illustrates higher levels of enzymes involved in ammonia assimilation, higher levels of glutamate dehydrogenase, glutamine synthetase, and two aminotransferases (Dembinski et al., 1991; reviewed in Below et al., 2004).

Because protein and starch are inversely related, carbon metabolism may also be altered by selection for grain protein. One aspect of carbon metabolism that was shown to differ between IHP and ILP was the level of ADP-glucose pyrophosphorylase, a key enzyme regulating starch biosynthesis

(Reggiani et al., 1985; Below et al., 2004). ILP contained higher levels of this enzyme than IHP, consistent with the fact that ILP accumulates more starch than IHP. ILP also channels more sucrose to the grain, while it simply accumulates in the stalk in IHP.

Protein QTL Studies

It is apparent from previous research that long-term selection for grain protein concentration has altered a number of related traits, and from this it may be hypothesized that many genes have also been affected (Moose et al., 2004). This hypothesis is supported by the progress of the strains after over 100 cycles of selection (Dudley and Lambert, 2004), which illustrates the quantitative nature of these traits. Utilizing populations created by crossing individuals from cycle 70 of IHP and ILP, or IHO and ILO (Dudley et al., 1977), several genetic mapping studies have identified quantitative trait loci (QTL) influencing protein (Goldman et al., 1993; Dijkhuizen et al., 1998; Dudley et al., 2004; Clark et al., 2006; Dudley et al., 2007) and oil concentrations (Clark et al., 2006). The results of these studies suggest the presence of many QTLs having small phenotypic effects, which is consistent with theoretical estimates for the number of effective genetic factors based on genetic variances, 102 to 178 genetic factors for protein and 14 to 69 factors for oil (Dudley and Lambert, 1991; Dudley et al., 2004).

Random mating the progeny of the IHP x ILP and IHO x ILO crosses was a strategy employed to reduce linkage disequilibrium that was predicted to increase following selection. These random-mated populations also offer advantages to genetic mapping experiments by increasing recombination and hence precision of associations between markers and QTL. Analysis of populations following seven generations of random mating (RM7) showed that this strategy was successful in reducing linkage disequilibrium (LD) and breaking up coupling-phase linkages among genes controlling protein concentration, as assessed in crosses of the RM7 population to two common testers (Dudley, 1994;

Dudley et al., 2004). Ten generations of random mating following the cross of IHO x ILO produced similar results (Clark et al., 2006).

Selection for grain protein concentration has altered the relative amount of the major seed storage proteins, the zeins (Doehlert and Lambert, 1991). Because zein protein is deficient in lysine and tryptophan, and also negatively correlated with yield, decreasing zein accumulation is one strategy for increasing both nutritional quality and yield. One goal, therefore, is to identify the number and identity of genes associated with reduced zein protein accumulation. The IRHP strain phenotype is characterized by a more desirable lower protein (and zein) concentration (7%). The rapid response of IRHP to reverse selection (**Figure 1a**) also suggests that reduced protein in IRHP may be controlled by a small number of genetic factors, which would facilitate introgression of the major IRHP gene into elite germplasm. For these reasons, IRHP1 is being used to further study the regulation of zein protein accumulation.

RESULTS

Chemical Analytical Procedures

N combustion and NIR are two possible methods for analyzing grain protein concentration of maize kernels, but NIR also allows for the simultaneous measurement of oil and starch concentrations, and relative moisture of the sample. Before using the NIR instrument to measure samples of unknown protein concentration, it is first necessary to create a calibration curve with samples having a broad range of known grain compositions. Approximately forty grain samples from the protein and oil strains were used to develop a NIR calibration for the DICKEY-John instrument in 2006. The samples ranged in protein from 4.4% to 28.74%, oil from 1.5% to 20.4%, and moisture from 9-10%. We did find that the percentage of moisture in the kernel affects grain composition estimates, with oil being particularly sensitive to moisture levels outside of the range from 9-10% that is typical for grain stored under ambient laboratory conditions. Thus, all samples were measured at 9-10% moisture, which was achieved by drying the sample following grinding, if necessary.

Approximately 60 ears from each of the IHP, IRHP, IRLP and IRLP2 populations grown in the 2006 field season, as well as the bulks of the Illinois Oil Strains, a total of 241 samples, were analyzed for grain protein concentration utilizing both N combustion and NIR. The absolute estimates of protein concentrations were typically lower using the NIR method, but showed a nearly perfect correlation of 0.99. The protein concentrations obtained by both methods were ranked and the ranked values correlated and plotted **(Figure 2).** The rank correlation coefficient between data obtained by both methods was 0.95, illustrating comparable results from NIR as N combustion, and the ability to identify the individual ears with extreme phenotypic values for grain protein. Similarly high rank correlations were observed between measurements of oil concentrations by nuclear magnetic resonance and the NIR instrument (D. Roberts, data not shown). Therefore, beginning in 2006, all strains of the Illinois Long-term Selection Experiment have been measured using NIR.

The estimation of protein and oil concentrations of unknown samples by NIR is most accurate when the composition of unknown samples lies within the range of those used to develop the calibration. The ability to accurately estimate protein and oil concentrations that exceed the range of the calibration is subject to potential errors associated with extrapolating beyond the known data set. This prediction error is expected to increase with samples where grain composition greatly exceeds the values in the calibration set. This is especially problematic as the rate of gain slows in IHP and IHO because small changes due to genetic effects may lie within the error associated with NIR measurements, making it difficult to detect significant changes in the means due to selection. Due to the continual deliberate selection of extreme values, it is not surprising that oil concentrations in IHO (22.4%) and protein concentrations in IHP (35.8%) have already exceeded the range of standards used to build the calibration. For these reasons, it may be necessary to re-calibrate the NIR instrument every few years using the widest possible range of samples. However, the currently high rank correlation and low degree of dispersion around the fitted relationship of the data is quite small, indicating a stronger confidence in the extrapolations. Nevertheless, a calibration every five or ten years is recommended.

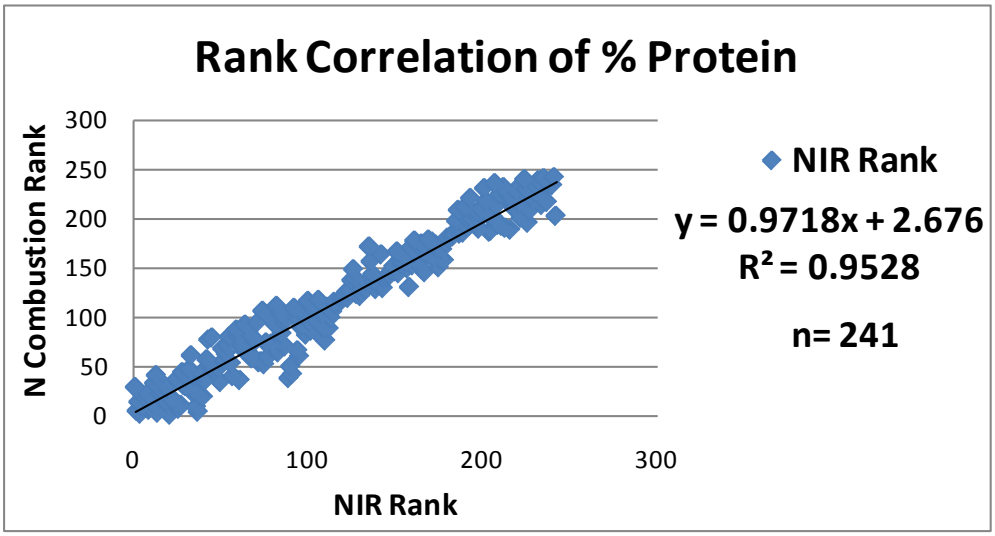


Figure 2. Rank correlation of protein concentrations estimated from 241 individuals of the IHP, IRHP, IRLP and IRLP2 populations grown in 2006, utilizing NIR and N combustion techniques.

Illinois Protein and Oil Strains

Protein means are plotted for IHP, ILP, IRHP, IRLP, IRLP2, IRHP2 and IRHP3 in **Figure 1 (a)** and oil means for IHO, ILO, IRHO, IRLO and ISHO in **Figure 1 (b)** as a function of year. This data is available for download at the following website: http://cropsci.illinois.edu/faculty/moose/lab/download.html. While we plot the means according to year, they have previously been plotted as a function of generation and cycle. Here, we define each of these terms as they are used with respect to the Illinois Long-term Selection Experiment. Although spanning a period of 115 years (1896-2010), the experiment was not conducted for six years within this period, including three years during World War II (1942-44), and crop failures due to extremely hot and dry conditions in 1933 and 1971 . The first year of the experiment was conducted in 1896 and represents both the initial source population and the first cycle of selection for IHP, ILP, IHO and ILO. A generation is defined as a year in which the experiment was conducted, and 2010 marks the 109th generation of the ILTSE. The term "cycle" is often used interchangeably with "generation," but cycle refers to the number of cycles of selection and is actually specific to each strain, rather than the experiment as a whole. Discrepancies between cycle and generation have arisen in all of the strains throughout the course of the experiment for varying reasons. For example, although the experiment is in its 109th generation, the original reverse strains have only undergone approximately 60 cycles of selection due to their later initiation in 1948. Additional sources of discrepancy arise when data is not collected for a particular strain in a year in which data was collected for the other strains. IHP and IHO, for instance, have only experienced 108 cycles of selection. Thus, while the number of selected cycles varies for each strain for different reasons, the number of generations remains the same, increasing every year the experiment is conducted. Here, we plot the means by year to illustrate generations in which the experiment was not conducted, as well as years in which a cycle of selection was skipped for a particular strain (**Figure 1**). Additionally, the generation and cycle numbers for all of

the strains are provided in **Table 1**. Overall, the experiment has conducted a cumulative total of 721 cycles of selection for grain composition.

	IHP	ILP	IRHP	IRLP	IRLP2	IRHP2	IRHP3	IHO	ILO	IRHO	IRLO	ISHO
Generation	109	109	109	109	109	109	109	109	109	109	109	109
Cycle	108	93	59	61	19	7	8	108	86	58	61	53

Table 1. Generation and cycle numbers for all strains of the Illinois Long-term Selection Experiment. While the experiment spans 115 years (1896-2010), it was only conducted in 109 years. Thus, the experiment is in its 109th generation. Discrepancies between the number of cycles of selection and generation have arisen when data for a particular strain is not collected during a year in which the experiment is conducted. The reverse strains exhibit lower cycle numbers because they were initiated later in the experiment, and ILP and ILO because they were discontinued. Data for IHP and IHO has been collected in every generation except for one.

One reason for creating IRHP2 and IRHP3 was to test the effect of nitrogen fertilizer application on protein concentration, which has not been previously determined in the IPS. We chose to examine the impact of nitrogen on these IHP-derived populations because it was expected from the variability in mean protein concentrations for IHP (**Figure 1**) that the IHP phenotype would be most sensitive to soil N levels. IRHP2 plants were grown for three consecutive years (2004 - 2006) without supplemental N in an N-responsive field (Uribelarrea et al., 2007) and IRHP3 plants were grown in adjacent plots supplemented with 200 kg N per hectare. The protein concentrations of the IRHP2 population were 2.6 to 4.5% lower than IRHP3 over the three years.

Novel Methods for Calculating Genetic Gain

Previous methods for estimating gain from selection do not take into account the proportion of gain that could be due to environmental factors. One method for estimating environmental variation is to grow control populations that are not subjected to selection for the trait of interest in each generation. This allows for estimation of the effects of environmental variability on phenotypes within the control strains, which can then be considered estimates of genetic responses to selection. Because

grain protein concentrations from the bulks of IHO and ILO ears were also obtained each year that oil was measured, and because selection for oil is not altering protein, IHO and ILO serve as control strains throughout the course of the experiment for the Illinois Protein Strains. Similarly, grain oil concentrations from the bulks of IHP and ILP ears were measured each year protein was measured, and selection for protein is not altering oil. Thus, IHP and ILP serve as controls for the Illinois Oil Strains.

One novel aspect of the graphs in **Figure 1** is the plotting of the protein means of IHO and ILO with the Illinois Protein Strains (**Fig. 1a**) and the plotting of the oil means of IHP and ILP with the Illinois Oil Strains (**Fig. 1b**), as depicted by dashed lines. Plotting the controls in this fashion allows for a visual representation of the extent of environmental variation over the last 109 generations of the experiment, and may also indicate the extent of genetic drift in these populations. IHO and ILO protein means show a slight positive trend since 1896, and range from 8-20% protein throughout the experiment, which illustrates the maximum degree of phenotypic change in related populations that have not been selected for grain protein concentration. The oil means of IHP and ILP allow for a similar estimation, but have only varied by 2-7% oil. In addition to providing an estimate of environmental variability on grain protein and oil concentrations, this data also suggests that protein is more sensitive to environmental fluctuations than oil. This is also exhibited by global peaks or depressions in protein concentration among all of the Illinois Protein Strains in certain years. For example, significant peaks in protein are evident in all of the protein strains in 1934 and 1990, suggesting that protein is sensitive to environmental factors and that these years exhibited favorable conditions. Due to the discontinuation of ILP and ILO strains, however, a new set of controls is needed. Here we use inbreds derived from generation 90 of the Illinois Protein Strains: IHP1, ILP1, IRHP1 and IRLP1. Each of these strains has been grown on a yearly basis since 2006 and protein concentrations were obtained by NIR. Since they are inbred, their protein concentrations are not predicted to vary significantly from year to year, and any variation is likely due to the environment. Estimates of phenotypic variation in the control inbreds can

then be used to estimate the variability in the IPS populations due to environmental factors, which may be useful in increasing the power of genetic selection within a given generation. We are currently generating inbred lines from the most recent cycles of the Illinois Oil Strains and they will be useful for estimating the effect of environment on oil concentration. Until then, we will continue to use oil concentrations of the inbred Illinois Protein Strains (IHP1 and ILP1) as controls.

Means of IHP, IRHP, IRLP, IRLP2, IRHP2 and IRHP3 populations and selected individuals from 2000-2009 are plotted in **Figure 3**. Additionally, beginning in 2006, IHP1 means are plotted with IHP, IRHP2 and IRHP3; ILP1 means are plotted with IRLP2; IRHP1 means are plotted with IRHP; IRLP1 means are plotted with IRLP. The primary goal of this analysis is to determine if progress is being made by selection, which is an indicator of the extent of genetic variability available for exploitation in the strains. Progress may be measured in each of the strains by comparing the slope of the population mean with the mean values for the IHP1 or ILP1 inbred lines grown in each generation. Neither IHP1 nor ILP1 exhibited significant deviations from a slope of zero for the best fit line from linear regression. As observed for the selection strains themselves throughout the experiment (**Fig.1**), IHP1 shows greater variability among years compared to ILP1.

IHP and IRLP both exhibit positive population slopes, indicating that protein is increasing in these strains and progress is being made by selection for high protein concentrations. The slopes of the selected individuals in each of these strains are also increasing, which also supports that progress is being made. The IRHP population means have actually increased slightly since the year 2000, despite selection for low protein, suggesting that progress may have ceased in IRHP. The means of selected IRHP individuals are also increasing, providing further support that selection is no longer altering protein concentration in IRHP. Following six cycles of selection for low protein concentrations in each of IRHP2 and IRHP3, the slopes of selected individuals and populations are slightly positive, indicating no response to selection for low protein in either strain.

Based on the last analysis (Dudley and Lambert, 2004), IRLP2 population means were not increasing. However, clear increases in population means have been observed since that time, and are especially notable beginning in 2007. The apparent progress in IRLP2 indicates that, despite having undergone 90 cycles of forward selection, ILP contained enough genetic variability to be exploited by selection to achieve progress in IRLP2. The ILP1 means are also plotted with IRLP2, and there is no significant change in protein concentration among the four generations this inbred line was measured, further supporting the conclusion that IRLP2 is showing progress from selection during this period.

Progress in the Illinois Oil Strains may also be determined by analyzing the population means. Means of IHO, IRHO, IRLO and ISHO populations and selected individuals are plotted in **Figure 4**. The population slope of IHO is actually negative, even though the slope of the selected individuals is positive. This may indicate that selection is no longer increasing oil in the IHO population. Oil concentration in the IRHO and ISHO populations are increasing slightly, although the slope of the selected individuals in IRHO is negative and only slightly positive in ISHO. With such small changes in the population means in these strains, it is not clear if progress is still being made in IRHO and ISHO. Of all the oil strains, the IRLO population shows the greatest positive slope, albeit very small. The slope of IRLO selected individuals, however, is much greater (6.5 times) than the IRLO population, and also much greater than any slope in the other oil strains. Thus, if progress is indeed being made in any of the oil strains, it is at a very slow rate.

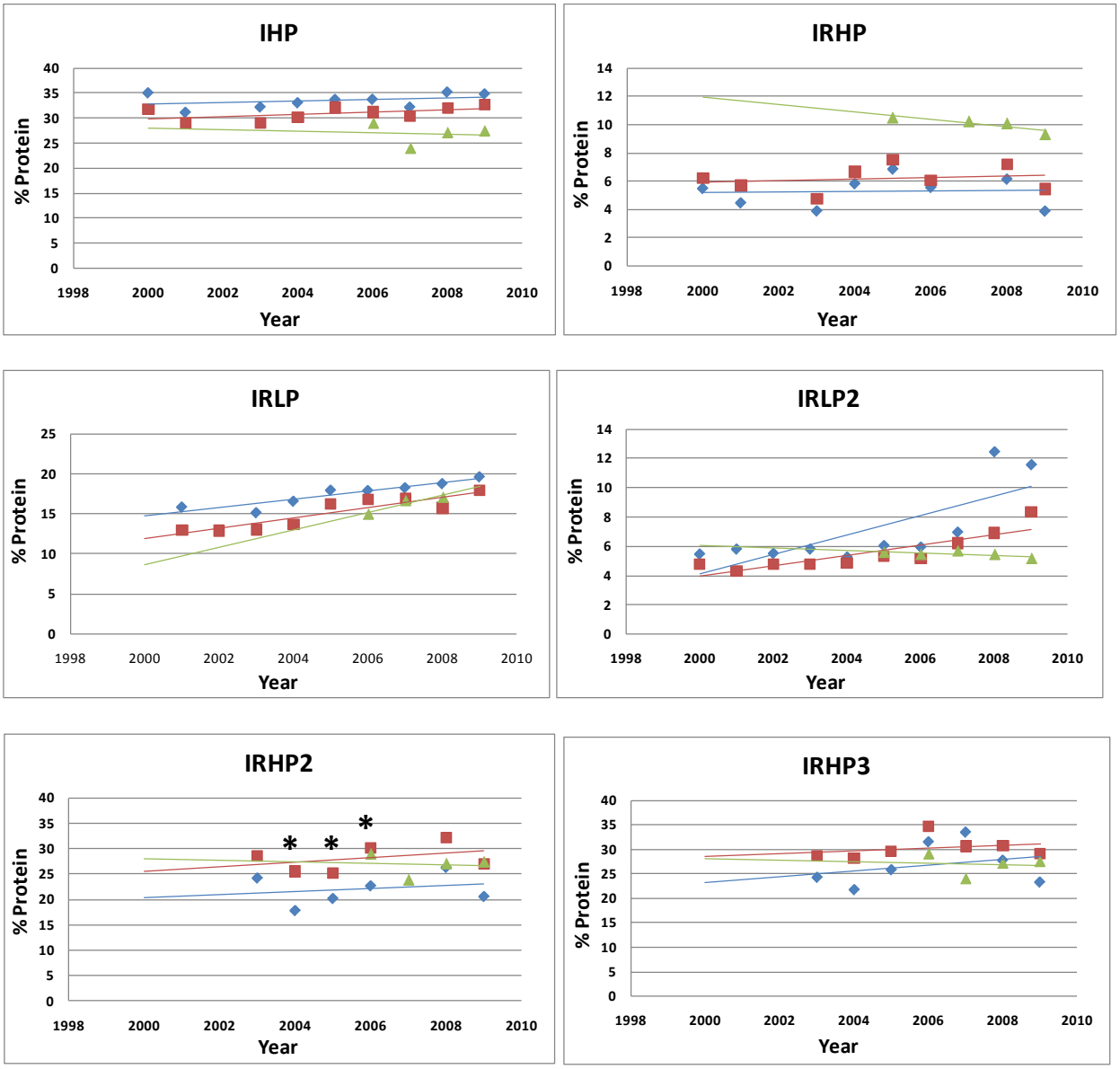


Figure 3. Mean protein concentrations in Illinois Protein Strains populations (red squares) and selected individuals for each generation (blue diamonds) from 2000 to 2009. Additionally, phenotypic means for the inbred line IHP1 (green triangles) are plotted with IHP, IRHP2 and IRHP3. Means for the ILP1 inbred line (green triangles) are plotted with IRLP2. Means for the IRHP1 and IRLP1 inbred lines are plotted with IRHP and IRLP, respectively. No additional N was applied in IRHP2 for years marked with asterisks.

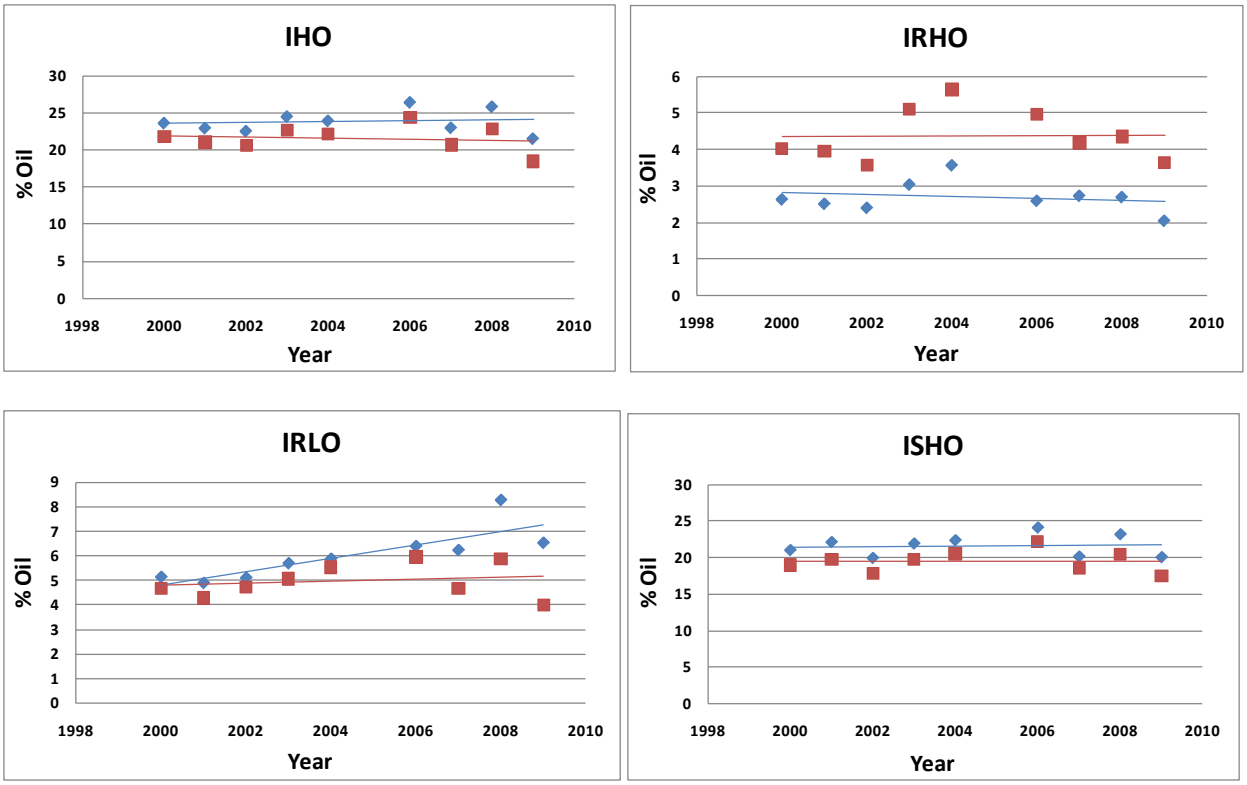


Figure 4. Illinois Oil Strains mean oil concentrations in populations (red squares) and selected individuals (blue diamonds) from 2000 to 2009.

IRHP may be Controlled by One or a Few Genes with Major Effects

The rapid decline of protein concentration in IRHP suggests that relatively few, and possibly only one major genetic factor may control the desirable low protein phenotype of IRHP. Therefore, IRHP provides an ideal genetic background for studying the regulatory mechanisms underlying protein concentration. In order to study the segregation of protein in IRHP, crosses were made between the IRHP1 and IHP1 inbred lines. However, genetic analysis of the progeny is complicated by the strong maternal effect on grain protein and zein concentrations, where the grain protein concentration of the progeny follows that of the maternal genotype. This maternal effect is well documented in maize, where it has been observed in the progeny of crosses between a number of genotypes, including crosses between IHP and ILP, as well as these genotypes with elite inbred varieties (Tsai et al., 1990). The maternal effect has not, however, been documented in crosses between IHP1 and IRHP1.

To study the maternal effect, F1 hybrids between IHP1 and IRHP1 were reciprocally backcrossed to each inbred parent. Zein protein was measured for individual kernels within these first-generation backcross (BC1) ears, and little variation is observed, which is consistent with a strong maternal effect uniformly influencing protein concentrations despite the genotypic segregation occurring within individual seeds. Furthermore, the mean protein concentrations of BC1 kernels on a given ear reflected the protein concentrations observed for the female parent in the cross, being highest in crosses to IHP1 ears, intermediate in crosses to the F1 hybrid ears, and lowest for IRHP1 ears.

The BC_1 seeds analyzed for zein were planted on a larger scale and the individuals self-pollinated to create two populations of BC_1S_1 ears: [IRHP1 x (IHP1 x IRHP1)]S_1 and [(IRHP1 x IHP1) x IRHP1]S_1. Selfing these individuals was done to control for the maternal effect and also to create a population segregating for IRHP1 and IHP1 alleles, where the progeny are either homozygous IRHP1 or heterozygous for IRHP1 and IHP1 at any given locus. Protein concentrations were measured on 78 individual ears from the first population and 37 individual ears from the second population. Regardless of genotype, both populations were characterized by nearly equal means (12.9% and 12.6%) and a similar bimodal distribution split at approximately 13% protein. Based on this information, the two populations were combined and analyzed jointly, which is plotted in a frequency histogram for protein concentration (**Figure 5)**. This larger population contained 51 individuals with protein concentrations less than 13%, with a mean of 10.6%. The remaining 64 individuals exhibited protein concentrations greater than 13% with a mean of 14.5%. These observations are consistent with equal segregation into two phenotypic classes of greater than or less than 13% protein (χ^2 test 1.47, P =0.22, df =1).

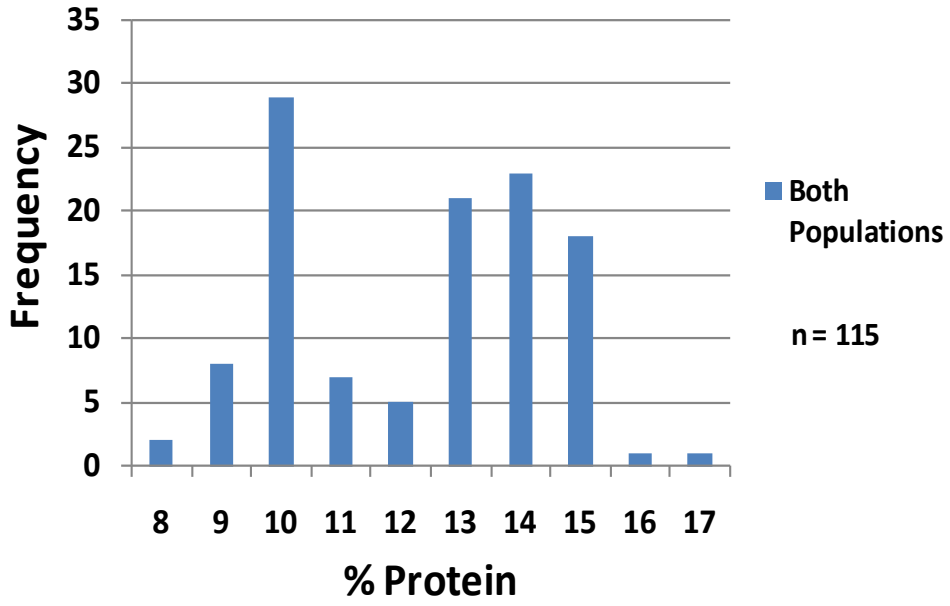


Figure 5. Protein concentrations of BC_1S_1 ears (n=115) from the cross of IRHP1 and IHP1 inbred lines.

Previous research investigating N metabolism in the IPS illustrates a hyper-accumulation of asparagine (Asn) in IHP (Dembinski et al., 1991; Lohaus et al., 1998). The primary route for Asn synthesis in plants is through the enzyme asparagine synthetase (AS), which has been reported to exhibit increased activity and mRNA expression (Lohaus et al., 1998). Conversely, the enzyme L-asparaginase (ASNase) plays an important role in release of stored ammonia from asparagine. Thus, altered regulation of genes encoding AS3 or ASNase could contribute to the IHP phenotype. The results from other studies in the Moose laboratory have shown that Asn accumulation in IHP is due to the combined up-regulation of RNA expression and protein accumulation for *ASPARAGINE SYNTHETASE3 (AS3)* and greatly reduced *ASNase* expression (Church, 2008). Markers that distinguished polymorphic promoter sequences in the *AS3* and *ASNase* alleles for IHP1 and IRHP1 were developed by Farag Ibraheem and available for genotyping the BC_1S_1 population created here.

These markers were used to test for significant associations between *AS3* and *ASNase* allelic variants and the protein concentration phenotypes in the BC_1S_1 population. If either one of these genes

was the cause of the observed bimodal distribution in this population, we would expect to see strong associations between genotypic and phenotypic data, which would be reflected by significant differences in protein concentration between the two genotypic classes (IHP1/IRHP1 heterozygotes and IRHP1/IRHP1 homozygotes). The mean grain protein concentrations for *ASNase* were not significantly different among the genotypic classes (α 0.05; p value 0.7485). However, the mean concentrations for *AS* was 13.39% for heterozygotes and 11.97% for homozygotes (α 0.05; p value 0.0011). These results indicate that although *ASNase* does not account for the bimodal distribution of grain protein concentrations in the BC_1S_1 population, *AS3* does appear to be associated with protein concentration, with the allele present in IHP1 contributing to increased protein in the IHP1/IRHP1 heterozygous individuals.

DISCUSSION

Analysis of the IPS indicates that protein concentrations in IHP, IRHP and IRLP are still increasing, suggesting that a limit to selection has not been reached in these strains. The protein concentration of IRLP2 had not shown any progress after ten cycles of selection based on the last analysis (Dudley and Lambert, 2004), but has increased in the last decade, with the most remarkable increases in 2006-2009. These results indicate that, despite ninety cycles of selection for low protein concentration in ILP, enough genetic variation existed in ILP, either remaining from Burr's White or novel variation due to mutation, to increase protein in IRLP2. The observed delay of approximately ten to fifteen cycles before the phenotypic reversal began was also apparent in IRHP and IRLP, and may be evident in IRHP2 and IRHP3 as well, as indicated by the lack of progress in these lines since their initiation in 2003. On the contrary, none of the Illinois Oil Strains exhibited gain from selection. One possibility for this trend, at least with respect to IHO, is that the oil concentrations of the samples lie outside the range of the standards used to build the calibration. Concentrations of the other strains, however, do lie within the range measurable by NIR, so lack of progress in these strains most likely indicates that a limit to selection has been reached. Since the Illinois Oil Strains have already been exploited for favorable oil genes in developing high oil varieties (Alexo High Oil, etc.), it is reasonable, therefore, to discontinue selection in these strains.

The use of control populations that are not changing for the trait under selection will provide a more powerful method for estimating the proportion of the phenotypic variation attributed to environmental and experimental variation. In the future, the inbred IPS and IOS will be methodologically planted with the ILTSE for this purpose. The results of growing RHP2 and RHP3 under low and high N have definitively shown the effects of N fertilizer on grain protein concentration, and this source of variation can be differentiated from the experimental variation and genetic drift observed in earlier generations of the experiment. Given the relatively small effect of N on protein concentration,

subsequent cycles of both IRHP2 and IRHP3 were conducted in plots receiving a typical level of fertilizer N (100 kg per hectare), and both new strains produced similar grain protein phenotypes. Because the IRHP2 and IRHP3 strains have yet to show a response to selection for low protein, they can effectively be considered replicates of IHP cycle 103 with the highest known grain protein concentrations (>30%), where the effect of N fertilizer appears to contribute at most 4-5% to total grain protein concentration. This is certainly less that the observed rate of gain in both IHP and IRLP since the use of supplemental N fertilizer began with generation 53. The use of N fertilizer also did not inhibit further decreases in either ILP or IRHP. Thus, although N fertilizer use may have contributed to continued response to selection for high protein, we conclude it has a minor impact compared to genetic changes.

The phenotypic results of the BC_1S_1 population between IHP and IRHP strongly support the hypothesis of few genes having major effects. The population was characterized by a bimodal distribution with approximately equal numbers of individuals belonging to each phenotypic class (>13% or <=13%). If the trait was controlled by a larger number of genetic factors, a more continuous distribution would be expected with fewer numbers of individuals belonging to more phentoypic classes. The results of the Chi-Square analysis indicate that the IHP1 allele of *asparagine synthetase (AS)* is correlated with higher protein concentration in IHP1/IRHP1 heterozygous individuals, but it does not appear to be responsible for a large percentage of phenotypic variation. The BC_1S_1 population analyzed here represents a good starting point for coarse genetic mapping studies, but a larger BC_1S_1 population consisting of 200+ individuals was generated in the summer of 2010 for additional fine-mapping. This information will be used to identify gene candidates associated with the more desirable reduced zein phenotype of IRHP, and also to facilitate marker-assisted introgression of the major IRHP1 gene(s) into other elite germplasm. It would be interesting to create a similar BC_1S_1 population between IHP1 and ILP1 to attempt to identify genes controlling protein in ILP.

Analysis of the IRHP population (**Figure 1a**) provides additional evidence supporting the theory of fewer genes having large phenotypic effects. It took only fifteen generations of reverse selection to revert nearly 50 generations of gain in IHP, illustrating the fastest rate of change in protein concentration in any of the strains throughout the experiment (**Figure 1a**). This observation is consistent with the trait being controlled by one major genetic factor with the possibility for other minor factors, at least with respect to the reversal of protein concentration in IRHP. Furthermore, at the time IRHP was initiated, genetic variability available for exploitation by selection was already reduced in IHP. Forty-eight cycles of forward selection for high protein concentration in IHP likely eliminated "low protein" alleles from the population, leaving behind fewer factors that could be exploited by selection for low protein in IRHP, which would be predicted to have larger effects on the phenotype. Additionally, marker analysis on all four IPS revealed a close relationship between IHP and IRHP. The close relationship between IHP and IRHP is consistent with IRHP being derived from IHP and indicates a low amount of genetic variation between the two strains despite the large phenotypic divergence (Mikkilineni and Rocheford, 2004).

MATERIALS AND METHODS

Illinois Long-term Selection Procedures

Selection procedures are outlined in Dudley and Lambert (2004), and sources for detailed information about the chemical analytical procedures are listed here as well. Briefly, 60-100 ears of each strain are analyzed for their respective traits, and a bulk of all ears for each strain is analyzed for the trait not under selection (e.g., a bulk of IHP was analyzed for oil concentration).

Phenotypic Trait Measurement

Approximately 60 ears from each of the IHP, IRHP, IRLP and IRLP2 populations grown in the 2006 field season, as well as the bulks of the Illinois Oil Strains, for a total of 241 samples. Approximately 100 seeds from each ear were ground to a fine flour and the flour was analyzed for grain protein concentration utilizing the DICKEY-john Instalab 600 near-infrared analyzer (NIR). Approximately 50 mg oven-dried flour was analyzed for total N using the Thermo Flash EA 1112 Series N combustion analyzer using the method of Dumas (Kirsten and Hesselius, 1983). 2007-2010 grain protein concentrations for the IPS and IOS were obtained using NIR. Grain protein concentrations of BC_1S_1 ears were measured using NIR.

Generation of the BC_1S_1 Populations

BC1 seed was available of [(IRHP1 x IHP1) x IRHP1] and [IRHP1 x (IHP1 x IRHP1)] for planting in the 2009 summer nursery. Twenty rows were planted with 20 kernels/ row for each genotype. Plants were grown to maturity and self-pollinated. Grain protein concentrations were obtained using NIR (see Phenotypic Trait Measurement).

DNA Isolation of BC_1S_1 population

Twelve discs of leaf tissue were hole-punched into 1.7 mL microcentrifuge tubes containing two steel balls. The PUREGENE method for DNA Isolation method (Quiagen) was used for DNA extraction following the manufacturer's protocol with modifications to accommodate samples collected in individual 1.7 mL microcentrifuge tubes instead of a 96-well plate. Samples were flash frozen in liquid N and stored at -80°C until isolation. Samples were ground using a GenoGrinder for 45 sec at 1500 rpm (setting 500 at 1x rate). 400 uL of a mixture of Cell Lysis buffer and RNAse were added to each tube (3 uL of 10mg/mL RNAse A/ 10 mL Cell Lysis Solution). Samples were immediately genoground to suspend powder in solution. Samples were incubated in a hot water bath at 65°C for 1 h. After incubation, samples were placed on ice and 100 uL of Protein Precipitation Solution was added. Samples were vortexed for 15 sec and incubated on ice for 30 min. Meanwhile, 600 uL of isopropanol were added to new 1.7 uL microcentrifuge tubes. After incubation, samples were centrifuged for 10 min at 14,000 rpm. 600 uL of the supernatant was added to the set of tubes containing isopropanol. Samples were incubated for 20 min at 37°C. After incubation, samples were centrifuged for 10 min at 14,000 rpm. Supernatant was poured off and allowed to drain from tubes by blotting and inverting on a paper towel. Pellets were washed 3 times with 70 % EtOH by adding 600 uL EtOH, vortexing for 10 sec, centrifuging for 5 min at 14,000 rpm, and then pouring off supernatant. Pellets were dried by incubating for 30 min, or until EtOH could no longer be detected, at 65°C. Samples were re-suspended in 100 uL distilled water and stored in -20°C.

Genetic Markers

Asparagine Synthetase (AS) SSR and *Asparaginase* CAPS markers were developed by Farag Ibraheem and Han Zhao, respectively, to identify and amplify polymorphisms between IHP1 and IRHP1 by PCR. The *AS*

primers anneal within the promoter region of *AS* on chromosome 1S and result in 2 bands for IHP1, 498 and 778 bp and one band for IRHP1, 557 bp. The primer sequences are as follows:

Forward: start 45110346bp, end 45110366bp; 5'-CTCAACTCATCGGCACAGACTTGCATC - 3'

Reverse: start 45110501bp, end 45110522bp; 5'- TCGAATTTATCCTTTCTACAACCCCAATC- 3'

The *ASNase* primers anneal the third exon of *ASNase* on chromosome 2 and the sequences are as follows:

Forward: start 5'-5060105bp, end 5060128bp; 5'-TCATGGAGTACAAGGGCCTGCC-3'

Reverse: start 3'-5060380bp; end 5060359bp-5' 3'-TTAAACACATGGCAATCGCAGGAT -5'.

The *ASNase* amplicons were cleaved with the Hpy188I restriction enzyme (NE Biolabs) and resulted in 2 bands of 73 bp and 241 bp for IHP1 and one band of 314 bp for IRHP1. 10x Standard Taq Reaction Buffer (NE Biolabs), dNTPs (Biorad), and taq polymerase (NE Biolabs) were used for PCR. The PCR amplification profile used an initial denaturation step of 94°C for 2m followed by 34 cycles of 94°C for 30s annealing at 53°C for 1m, and extension at 72°C for 1m.

WORKS CITED

Below, F.E., Seebauer, J.R, Uribelarrea, M., Schneerman, M.C. and Moose, S.P. (2004) Physiological changes accompanying long-term selection for grain protein in maize. *Plant Breeding Reviews* 24 (1): 133-151.

Church, J.B. (2008) Functional genomics of nitrogen use in maize. Ph.D. dissertation, University of Illinois.

Clark, D., Dudley, J.W., Rocheford., T.R. and LeDeaux, J. (2006) Genetic Analysis of Corn Kernel Chemical Composition in the Random Mated 10 Generation of the Cross of Generations 70 of IHO x ILO. *Crop Sciences.* 46: 807-819.

Dembinski, E., Rafalski, A., Wisniewska, I. (1991) Effect of long-term selection for high and low protein content on the metabolism of amino acids and carbohydrates in maize kernel. *Plant Physiology and Biochemistry.* 29 (6): 549-557.

Dijkhuizen, A., Dudley, J.W., Rocheford, T.R., Haken, A.E. and Eckhoff, S.R. (1998) Near-infrared reflectance correlated to 100-g wet-milling analysis in maize. *Cereal Chem*. 75: 266-270.

Dudley, J.W. (1977) Seventy-six generations of selection for oil and protein percentage in maize. *Proc. Intl. Conf. on Quantitative Genetics.* Ed. E. Pollak, O. Kempthorne, and T.B. Bailey, Jr. Ames, Iowa State University Press, 1977. 459-473.

Dudley, J.W. and Lambert, R.J. (1992) 90-Generations of Selection for Oil and Protein in Maize. Maydica 37: 81-87.

Dudley, J. (1994) Linkage disequilibrium in crosses between Illinois maize strains divergently selected for protein percentage. *Theor Appl Genet.* 87: 1016-1020.

Dudley, J.W., Dijkhuizen, A., Paul, C., Coates, S.T., and Rocheford, T.R. (2004) Effects of random mating on marker-QTL associations in the cross of the Illinois High Protein x Illinois Low Protein maize strains. *Crop Sci*. 44: 1419-1428.

Dudley, J.W. and Lambert, R.J. (2004) 100 generations of selection for oil and protein in corn. *Plant Breeding Reviews* 24 (1): 79-110.

Dudley, J.W., Clark, D., Rocheford, T.R. and LeDeaux, J.R. (2007) Genetic analysis of corn kernel chemical composition in the random-mated 7 generation of the cross of generations 70 of IHP x ILP. *Crop Sci.* 47: 45-57.

Goldman, I.L., Rocheford, T.R. and Dudley, J.W. (1993) Quantitative trait loci influencing protein and starch concentration in the Illinois Long Term Selection maize strains. *Theor. Appl. Genet*. 87: 217-224.

Hopkins, C.G., Smith, L.H. and East, E.M. (1903) The structure of the corn kernel and the composition of its different parts. Univ. Illinois Agricultural Experiment Station Bulletin 87.

Kirsten, W.J. and Hesselius, G.U. (1983) Rapid, automatic high capacity Dumas determination of nitrogen. *Microchem. J.* 28: 529-547.

Laurie, C., Chasalow, S., LeDeaux, J., McCarroll, R., Bush, D., Hauge, B., Lai, C., Clark, D., Rocheford, T.R., Dudley, J. (2004) The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics.* 168: 2141-2155.

Mikkilineni, V. and Rocheford, T.R. (2004) RFLP variant frequency differences among Illinois long-term selection protein strains. Plant Breeding Reviews 24: 111-132.

Moose S.P., Dudley J.W., Rocheford, T.R. (2004) Maize selection passes the century mark: a unique resource for 21st century genomics. Trends in Plant Science 9: 358-364.

Reggiani, R., Soave, C., Di Fonzo,N., Gentinetta, E., and Salamini, F. (1985) Factors affecting starch and protein content in developing endosperms of high and low protein strains of maize. *Genet. Agr.* 39: 221-232.

Smith, LH. (1908) Ten generations of corn breeding. Ill. Agr. Exp. Sta. Bul. 128: 454-575. Reprinted as p. 65-94. In: J.W. Dudley (ed.) 1974. Seventy generations of selection for oil and protein in maize. Crop Sci. Soc. Am., Madison, WI.

Tsai, C.L., Dweikat, I. and Tsai, C.Y. (1990) Effects of source supply and sink demand on the carbon and nitrogen ratio in maize kernels. *Maydica* 35: 391-397.

Uribelarrea, M., Below, F.E. and Moose, S.P. (2004). Grain composition and productivity of maize hybrids derived from the Illinois Protein Strains in response to variable nitrogen supply. *Crop Science* 44: 1593-1600.

Uribelarrea, M., Moose, S.P., and Below, F.E. (2007) Divergent selection for grain protein affects nitrogen use efficiency in maize hybrids. *Field Crops Res*. 100: 82-90.

Woodworth, C.M., Leng, E.R. and Jugenheimer, R.W. (1952) Fifty generations of selection for oil and protein in corn. *Agron. J.* 44: 60-65. Reprinted as p. 121-132. In. J.W. Dudley (ed.) 1974. Seventy generations of selection for oil and protein in maize. Crop Sci. Soc. Am., Madison, WI.

CHAPTER 2

FLOURY2-RFP REPORTER LINES TO STUDY ZEIN GENE REGULATION

ABSTRACT

Long-term divergent selection for grain protein concentration has produced populations with the known phenotypic extremes for this trait, by dramatically altering the accumulation of the 19-kD and 22-kD α-zeins. Known regulators of α-zein protein accumulation are *OPAQUE2* (*O2*), the *Prolamin-box Binding Factor* (*PBF*), and factors influencing the folding of zeins into endosperm protein bodies. RNA analysis and measuring protein abundance are two effective approaches for studying the regulation of zein expression, but they are also expensive, destructive and laborious. Furthermore, study of individual zein genes is complicated by their high copy number and sequence similarity. An alternative inexpensive and nondestructive approach to investigate the regulation of zein expression is the use of transgenic *Floury2-mRFP* reporter lines (Dave Jackson's lab at Cold Spring Harbor), where the expression of the readily visible monomeric red fluorescent protein (mRFP) is controlled by the genomic sequences encoding the *Floury2* α-zein. The *FL2-mRFP* transgene has been introgressed into inbred lines derived from the four Illinois Protein Strains (IPS), as well as the reference inbred B73. We found that *FL2-mRFP* expression not only correlates with grain protein concentration, but also follows known patterns of zein accumulation throughout development. At all developmental stages, RFP expression was strongest in Illinois High Protein (IHP), the lowest in Illinois Low Protein (ILP) and intermediate in Illinois Reverse High Protein (IRHP), Illinois Reverse Low Protein (IRLP), and B73. By crossing *FL2-mRFP* to an *o2* mutant introgressed into IHP, we show that its expression is strongly activated by *O2*, illustrating that the *FL2-mRFP* transgene is regulated in the same manner as endogenous α-zein genes. Future experiments will use the *FL2-mRFP* transgene as a tool for identifying regulators influencing protein concentration in ongoing genetic mapping studies.

INTRODUCTION

Zein Genes and Zein Gene Expression

Cereals are an extremely important component of the human diet. Maize, wheat and rice provide over half of the world population's caloric intake (Godwin et al 2009). This proportion increases to 65 percent if we consider calories consumed from the animal products derived from livestock fed cereal grain (Godwin et al 2009). Despite meeting caloric needs, cereals are regarded as nutritionally poor because, although their protein content satisfies adult daily requirements, they are deficient in several amino acids essential to human and livestock consumption. For example, while the Food and Agriculture Organization recommends 5.5% of lysine for human nutrition, maize proteins contain less than 3% (Huang et al. 2004). Maize proteins are also deficient in tryptophan, threonine and methionine. For this reason, there has been a great effort to increase the nutritional composition of maize protein with a goal of creating Quality Protein Maize (QPM).

During maize seed development, nitrogen from the plant is mobilized into the kernel where it is used to synthesize endosperm storage proteins that become the nitrogen source for the developing seedling. The majority of storage proteins assembles into protein bodies and accumulates in the subaleurone and starchy endosperm; they constitute 70% of total kernel protein and 85% of total kernel dry weight at maturity (Tsai, 1989). Of the endosperm storage proteins, approximately 50 -60% belong to the alcohol-soluble, proline- and glutamine-rich group of storage proteins, the prolamins, collectively termed zeins in maize. Therefore, the zeins comprise the largest proportion of proteins in the kernel and probably influence grain nutrition the most. As a result, their regulation has been the focus of much research over the past several decades in achieving QPM. Prolamins of other cereals are called kafirins (sorghum), gliadins (wheat), hordeins (barley) and oryzins (rice).

Zein proteins can be classified into four gene families based on differential solubility and polypeptide compositions: the 19 and 22 kD α-zeins, 14 kD β-zeins, 27 and 16 kD γ-zeins and 10 and 18

kD δ-zeins (Esen, 1986). However, since the α-zeins account for approximately 70% of the total prolamin fraction in the endosperm, it was the first family to be described and is well-characterized. The α-zein genes are encoded by two large multigene families containing an estimated 110 to 130 gene members and map to five regions in the genome: 4L, 4S, 7S, 10L and near the centromere of chromosome 1 (Viotti et al., 1979). The α-zein genes can be further resolved into gene subfamilies according to sequence homology,(Song and Messing, 2002) and consist of the z1A, z1B, z1C and z1D subfamilies (Shewry and Casey, 1999). The z1A, z1B and z1D subfamilies share a molecular weight ranging from 19- 21 kD, and are collectively referred to as the 19kD α-zeins. The z1C subfamily ranges from 22- 23 kD and is referred to as the 22kD α-zeins. The observed range in molecular weights occurs due to internal insertions and deletions in individual copies of duplicated genes within each subfamily (Heidecker et al., 1991). For example, 23 tandemly duplicated genes encoding the 22 kD α-zeins have been reported, with 22 gene copies tandemly arranged within 168 kb on chromosome 4S and the twenty-third gene copy located approximately 20cM away from the cluster and corresponding to the *Floury2* locus (Song et al., 2001; Coleman et al., 1995).

Both the 22 kD and 19 kD zein clusters have been sequenced in B73 and the 22 kD cluster has also been sequenced in BSSS53 (Song and Messing, 2001), facilitating more detailed studies of zein gene expression. The results of these studies indicate that the zeins are by far the most highly expressed genes in the endosperm (Woo et al 2001). The 19 and 22 kD α-zein genes are expressed in a coordinated fashion, beginning around 12 DAP, peaking around 16 DAP and continuing throughout development, with RNA expression patterns closely matching protein accumulation (Woo et al., 2001 Langridge et al., 1982). However, despite a large copy number, only a fraction of α-zein gene coding sequences produce transcripts in any given genotype, suggesting the presence of pseudogenes interspersed with functional genes (Marks et al., 1995; Woo et al., 2001). For example, of the twenty-three 22kD α-zein genes present in inbred line BSSS53, only seven genes were expressed, including the

floury2 allele (Song, 2001). In the B73 inbred, only six of 16 22 kD α-zein genes are expressed (Song and Messing, 2003; Feng et al., 2009). The 19kD α-zeins followed similar expression patterns, where 8 of 12 were expressed in the z1A gene family, 2 of 8 in the z1B family, and 2 of 5 in the z1D family in B73 (Feng et al., 2009). Furthermore, variation was also observed in the level of expression of the zeins among genotypes.

Additional results from gene expression studies indicate differences in the number of zein gene family members shared by the lines. For example, B73 and BSSS53 were shown to share 16 non-allelic and 11 allelic α-zeins, B73 and W22 shared 8 non-allelic and 14 allelic genes, and BSSS53 and W22 18 non-allelic and 10 allelic genes, thereby revealing dramatic expressional divergence of the α-zeins (Feng et al., 2009). Nevertheless, approximately the same number of genes was expressed in each inbred line, and several major haplotypes were discovered. Feng et al. also reported the presence of oscillating patterns of zein expression. While z1A and z1B expression levels fluctuated in a 4 day interval, z1C and z1D levels fluctuated in a 6 day interval (2009). Variation was also seen in peak expression times, where z1A peaked at 18 DAP, z1B at 22 DAP, and z1C and z1D at 24 DAP. These results illustrate the dynamic nature of zein genes. They are rapidly evolving in terms of copy number, relative chromosomal position, and expression. Based on these observations, one possible mechanism for altered protein concentration may be changes in the number, organization or regulation of the α-zeins.

Known Regulators of Zein Gene Expression

Due to their abundance and coordinated regulation, the zeins have been a model system for studying regulation of a multi-gene family. Known regulators of the zeins include *Opaque2 (O2),* a bZIP transcription factor, that binds a highly conserved sequence element containing an ACGT core 300-bp upstream of the start site of 22 kD α-zeins (Schmidt et al., 1992). The discovery of mutations in *O2* which led to reduced zein accumulation and increased levels of essential amino acids was a significant

step towards achieving QPM (Dalby and Tsai, 1975). For example, the *o2* mutation caused a 45 to 65% reduction in zein accumulation in the inbred variety W64A (Hunter et al., 2002) and a 700% increase in soluble lysine (Azevedo et al., 2003). Nevertheless, the *o2* phenotype is plagued by a softer endosperm, making kernels susceptible to insect herbivory, mechanical damage during harvest, and subsequent microbial infection, all of which reduce yield. Additional mutations were discovered, including *o7* and *o6,* as well as the *fl2* and *fl3* mutations, but none circumvented yield reduction, spurring the use of modifiers, such as *sugary2* to suppress the opaque phenotypes. Therefore, alternative approaches are sought for decreasing zein synthesis and discovery of additional regulators will greatly assist in strategies for achieving QPM.

Another known regulator of the zeins is the *Prolamin-box Binding Factor* (*PBF*), a Dof zinc finger DNA binding protein, which binds a highly conserved sequence element (the prolamin box, 5'-TGTAAAG-3') found in the promoters of all zein genes (Carbajosa et al., 1997) as well as prolamin genes of barley, wheat, oats and sorghum (Forde et al., 1985). Both *O2* and *PBF* expression profiles show complementary accumulation beginning 10 DAP, 1-2 days before the synthesis of zein proteins, which is consistent with their role in regulating the zein expression. Additionally, the *O2* binding site is located only twenty nucleotides downstream of the prolamin box, and the PBF and O2 proteins have been shown to interact both in vitro and later in expression assays within rice and wheat endosperm culture cells, where they may act singly or additively in driving expression of zein genes (Vicente-Carbajosa et al., 1997; Hwang et al., 2004). However, neither *O2* nor *PBF* appear as significant factors affecting grain protein concentration in QTL studies, suggesting the possibility for other key regulators.

Post translational mechanisms have also been shown to regulate the zeins, including factors influencing the folding of zeins into protein bodies (Bagga et al., 1997). Protein bodies originate within the rough endoplasmic reticulum (Larkins and Hurkman, 1978) and are the site of zein synthesis by polyribosomes attached to the exterior of the protein body membrane (Burr and Burr, 1976). Strong

interactions between the, β- and γ- zeins with the α-zeins suggests that the 15 kD β- and 16 kD may play a role in incorporating the α-zeins into protein bodies (Kim et al., 2002). The *Floury1* gene has also been shown to regulate protein body formation; *Floury1* encodes a transmembrane protein of zein protein bodies and is shown to interact with 22 kD α-zeins by targeting them to both the periphery of the protein body where γ- zeins accumulate and the central regions of the body (Holding, et al., 2007). Another class of factors influencing protein body assembly are chaperone proteins such as the binding immunoglobulin protein (BiP), whose biological function is to mediate protein folding and targeting to the protein body (Fontes et al., 1991).

Zein gene expression is also subject to what is known as the maternal effect, where the grain and zein protein concentration of the progeny follows that of the maternal genotype. This observation is well documented in maize, where it has been observed in the progeny of crosses between a number of genotypes, including reciprocal crosses between IHP and ILP, as well as these genotypes with elite inbred varieties (Reggiani et al., 1985; Tsai et al., 1990; Tsai and Tsai, 1990). There are three possible causes of the maternal effect: genomic imprinting, dosage effect and the nutritional status of the plant. Genomic imprinting is a type of epigenetic modification causing differential expression of a gene depending on the sex of the parent that transmits it. Two types of imprinting exist: allele imprinting, where only alleles from a certain genetic background are affected by parent-of-origin-specific gene expression, and locus imprinting, where all known alleles from different backgrounds are under parent-of-origin control (Gehring et all, 2004). Only known to occur in the endosperm tissue in plants (Gehring et al., 2004), it is hypothesized that the evolutionary role of maternal imprinting in the endosperm might be to maintain control of gene expression affecting kernel growth and development because of the dependency of kernel development on the maternal organ, the cob (Alleman and Doctor, 2000). The α-zein genes are maternally imprinted at the DNA level by means of differential methylation patterns, depending on the parental heritage. It is proposed that methylation of paternal alleles inhibits zein gene

expression by altering the chromatin structure of certain zein genes (Lund et al., 1995; Lauria et al., 2004; for review see Bird and Wolffe, 1999). Maternal alleles are demethylated, and thus de-repressed. Parental imprinting is also known to regulate expression of an allele of the *dzr1* locus, a posttranscriptional regulator of the 10 kD δ-zeins (Chaudhuri and Messing, 1994).

Dosage effect arises from the double fertilization of the endosperm, where one male gametophyte fertilizes the egg cell to form a diploid embryo and the other male gametophyte fuses with the two haploid polar nuclei to form a triploid endosperm consisting of two maternal and one paternal genomes (2m:1p). Reggiani et al. tested whether total protein and zein protein concentrations in the endosperm exhibited the dosage effect by making reciprocal crosses between IHP and ILP (1985). Kernels with either two (IHP x ILP) or three (IHP x IHP) doses of the IHP genome contained similar concentrations of total (approximately 21%) and zein (approximately 15%) protein. Similarly, kernels with either two (ILP x IHP) or three (ILP x ILP) doses of the ILP genome contained approximately equal concentrations of total (4-5%) and zein (1-2%) proteins. The results of this study indicate that protein concentration in the endosperm is not strictly dependent on the endosperm genotype (ie. the dosage), but rather the genotype of the maternal parent.

Differences in grain protein concentration have been attributed to the nutritional status of the vegetative source tissues, which may lead to changes in N supply to the grain, and/or changes in the capacity of the grain to accumulate N, and illustrate the influence of source and sink regulation. Genetic variability in the nutrient status of source tissues has been shown to influence seed nitrogen/protein concentrations in many plants, including maize (Wyss, et al., 1991). Selection for grain protein concentration in the Illinois Protein Strains has led to physiological changes affecting nitrogen metabolism, where IHP plants demonstrate elevated N uptake, N assimilation by seedling leaves, and N remobilization from source to seed sink tissues compared to ILP plants (Below et al., 2004). Therefore, it

is reasonable to hypothesize that these physiological differences might play a role in regulating α-gene expression within the kernel.

Characterization of Zein Genes and their Expression in the Illinois Protein Strains

Changes in grain protein concentration in the Illinois Protein Strains result from changes in α-zein protein accumulation, and have been observed since 1922 (Showalter and Carr, 1922). One hypothesis for the divergence of α-zein protein accumulation among the Illinois Protein Strains is differences in zein gene expression. Initial gene expression studies by Wrage revealed greater expression of the 19 kD α-zeins in IHP compared to ILP and B73 throughout seed development (2005). IHP also exhibited greater expression of the 22 kD zein genes than ILP and B73 at 15 DAP, with B73 exhibiting intermediate expression (Wrage, 2005).

Additional expressional profiling/microarray experiments using inbred lines derived from the Illinois Protein Strains (Uribelarrea et al., 2004) illustrated a ten-fold higher expression of both 19 kD and 22 kD α-zeins in IHP, IRHP and IRLP when compared to ILP (Martha Schneerman and Han Zhao, unpublished). Additionally, when probed with sequences specific to individual 19 kD or 22 kD genes, all 19 kD genes and all 22 kD genes showed consistent patterns of expression, suggesting that selection has altered a trans-acting mechanism that coordinately regulates all 19 and 22 kD α-zeins. *O2*, *PBF*, and BiP were also present on the microarray, but only the expressional profile of PBF followed α-zein protein accumulation. Thus, α-zein gene expression in the Illinois Protein Strains is coordinately regulated, most likely by a trans-acting mechanism other than *O2* or *BiP*, and possibly involving *PBF*.

Use of zein promoter-reporter genes to investigate zein gene expression in the Illinois Protein Strains

Although some mechanisms of zein gene regulation have been discovered, it is apparent that additional mechanisms underlying the coordinated expression of the 19 and 22 kD α-zein genes are

important, and their discovery should assist in creating QPM. However, approaches for studying individual zein gene expression are complicated by the high number and sequence similarity of the α-zein genes and the presence of pseudogenes interspersed with functional genes that have arisen from duplication events. Also confounding analysis is the quantitative inheritance of grain protein concentration, where greater than 100 genetic factors have been estimated to contribute to the protein concentration differences between IHP and ILP (Dudley and Lambert, 1991; Dudley et al., 2004).

One approach for tracking expression of individual zein genes is through the use of reporter genes. When fused to a zein gene promoter, the expression of the reporter gene can be measured in a rapid and quantitative way to estimate expression of the zein gene of interest. One of the first examples of this approach in maize was the fusion of the *E. coli* β—glucoronidase (GUS) reporter gene to a 27 kD γ-zein promoter (Russell and Fromm, 1997). GUS activity was shown to be proportional to transgene RNA expression, and the presence of the transgene did not affect endogenous zein gene expression. However, due to the necessary destruction of tissue sample to conduct the GUS fluorometric assay , GUS has limitations that prevent its use in certain applications, such as when the tissue must be saved.

In the last decade fluorescent proteins have increasingly been used in plant biotechnology, including transformation, gene expression and protein localization studies. The fluorescent protein reporter lines employed here are a part of a series of transgenic reporter maize lines that were created as part of a joint project conducted by Dave Jackson's laboratory at Cold Spring Harbor, Anne Sylvester's laboratory at the University of Wyoming, and the Plant Transformation Center at Iowa State University (Mohanty et al., 2009). With a goal of fusing fluorescent reporter genes to a set of gene sequences whose encoded proteins represent comprehensive localization to different cellular compartments, the reporter lines used here were originally created to characterize protein body localization and expression in maize (see http://maiaze.tigr.org/cellgenomics/index.shtml). Derived from the coral reef species,

Discosoma, a monomeric version of the tetrameric DsRed protein termed mRFP1 was fused to the C-terminus of the 22 kD α-zein gene, *Floury2* (Campbell et al., 2002).

To preserve proper tissue and temporal regulation of the *Floury2*-mRFP1 fusion protein, the construct includes an intact α-zein gene driven by native flanking regulatory elements, including approximately 2000-bp of the *Floury2* gene promoter and 1000-bp of 3' sequence. Inclusion of the native regulatory sequences may allow these lines to provide information about epigenetic regulatory pathways that influence gene transcription by modifying chromatin structure instead of DNA sequence. It may also allow for study of possible post-transcriptional mechanisms of regulation, including zein mRNA processing, localization and turnover. Thus, the mRFP reporter lines are a useful tool for studying regulation of α-zein gene expression.

When transformed into a HI-II background, initial analysis of the mRFP kernels by Campbell revealed the ability of the mRFP protein to be visualized under white light, the kernels appearing pink in color (2003). This property can most likely be attributed to the high level of expression of the *Floury2* gene, the relatively high stability of zein proteins, and the use of the monomeric RFP that does not require multimerization to emit fluorescence. *Floury2* has been shown to account for approximately 18-20% of all 22 kD α-zein expression in inbred variety B73 with only one gene expressed more (Song and Messing, 2003; Feng, et al., 2009). The ability to visualize *FL2-mRFP* expression under white light allows for the quick non-destructive monitoring of *Floury2* gene expression during seed development, which is not possible using other reporters, such as GUS, that require destruction of the tissue sample. Furthermore, since the α-zeins are coordinately regulated, *FL2-mRFP* expression will allow for tracking of not only 22 kD α-zein gene expression, but also 19 kD gene expression.

Since changes in α-zein gene expression underlie the dramatic divergence of protein concentrations in the Illinois Protein Strains, the IPS provide ideal genetic material for studying α-zein gene regulation. Therefore, one goal of this research is to introgress the *FL2-mRFP* transgene into the

four IPS and inbred variety B73. Initial crosses between the reporter lines and the IPS were done by Salas, and later generations have since been created (2008). Analysis of *FL2-mRFP* expression in BC1, BC3 and BC5 ears provides information about the genetic inheritance of the *FL2-mRFP* transgene, its response to genetic background and its expression throughout development. Additionally, the *FL2-mRFP* reporter lines are used to study the maternal effect on zein gene expression and protein accumulation, protein body formation and as a phenotype in quantitative trait mapping studies.

RESULTS

Genetic Inheritance of the *FL2-mRFP* Transgene when Crossed to the Illinois Protein Strains

The *FL2-mRFP* transgene was transformed by the *Agrobacterium* method into the Hi-II genetic background, a hybrid between inbred lines derived from crosses of A188 and B73, and the regenerated plants were backcrossed to the B73 inbred line before being provided to our lab. Seeds from three independent transformation events (47, 52 and 172) were used as the parents in crosses to inbreds derived from the IPS (Uribelarrea, et al. 2004): IHP1, ILP1, IRHP1, IRLP1 and inbred variety B73 (Salas, 2008). The most recent ears generated were BC5 ears (IPS) and BC6 ears (B73). However, photographs were taken of ears at various stages throughout the backcrossing, including BC1, BC3 and BC5, in order to document patterns of *FL2-mRFP* expression throughout the backcrossing process. Here, BC1 and BC5 ears are analyzed for clues revealing details about *FL2-mRFP* inheritance when crossed to the IPS.

Originating from the variety, Burr's White, the Illinois Protein Strains produce only white seeds. However, the parents of Hi-II, contain white kernels (A188) and yellow kernels (B73). Therefore, initial crosses to the *FL2-mRFP* lines produced ears containing seeds with varying degrees of yellow and white color, reflecting the continued segregation of kernel color from the original Hi-II genotype in these early crosses. Upon further backcrossing, however, less variation is observed in the yellow-white kernel color in all of the genotypes, indicating a decrease in segregation for this trait with increased backcrossing. **Figure 6** shows BC1 ears and **Figure 7** shows BC5 ears, where the *FL2-mRFP* transgene is transmitted through the female parent. Ears created by reciprocal crosses for both BC1 and BC5 ears produced similar results and are not shown here. BC1 ears illustrate more variation in yellow-white kernel color than BC5 ears.

Both BC1 and BC5 ears exhibit *FL2-mRFP* expression, as detected by the intensity of pink coloration of the kernels, and this coloration is consistent with the relative zein accumulation of the genotype into which it is backcrossed. IHP1 ears contain the most zein protein and show the highest

FL2-mRFP expression, as indicated by the dark pink coloration of the kernels. ILP1 ears contain the least zein and show the lowest *FL2-mRFP* expression, as indicated by the light pink coloration of the kernels. IRHP1 and IRLP1 contain intermediate levels of zein protein, and illustrate intermediate levels of mRFP expression. B73 also showed intermediate levels of expression, but quite lower than IHP. The coloration of the BC5 ears is expected to be more closely correlated to the protein concentrations of the inbred-derived Illinois Protein Strains since these ears contain a larger proportion of the recurrent parent genome. While BC5 kernels contain an average of 98% of the recurrent parent genome, BC1 kernels only contain an average of 75% of the recurrent parent genome. Therefore, BC1 ears still contain an average of 25% of the (Hi-II x B73) background and are likely still segregating for genes controlling protein concentration.

Regardless of genetic background, stage of backcrossing and transgenic event, approximately half of the kernels on an ear are white or yellow and the other half exhibits the red-to-pink coloration associated with the *FL2-mRFP* expression. This 1:1 ratio of white/yellow to pink kernels indicates that each transgene is inherited as a single dominant genetic locus. Furthermore, for all ears obtained, no variation in *FL2-mRFP* expression is observed among kernels within a single ear. This observation might be expected in later backcrosses where the kernels are no longer segregating, but is not expected in BC1 ears where kernels are still segregating. This data indicates that there is little genetic segregation within the transgene donor parents for factors regulating *FL2-mRFP* expression, and that whatever factors are regulating zein accumulation, influence the entire ear. Finally, *FL2-mRFP* expression is consistent among ears of differing transgenic events, despite having integrated in different locations in the genome due to random insertion by the *Agrobacterium* method of transformation. This indicates that the location of the transgene does not appear to affect its regulation. Nonetheless, all three transgenic events were backcrossed further into the four IPS and B73, if only to ensure stable behavior for at least one reporter line. Analysis of BC4 *FL2-mRFP* reporter lines, however, revealed no differences in *FL2-mRFP* expression

due to transgenic event in any of the genotypes (data not shown). For this reason, only event 52 was further backcrossed in the four IPS. All three transgenic events were fully backcrossed to the BC6 stage in B73, but no differences were observed, further supporting that transgenic event does not alter *FL2-mRFP* expression.

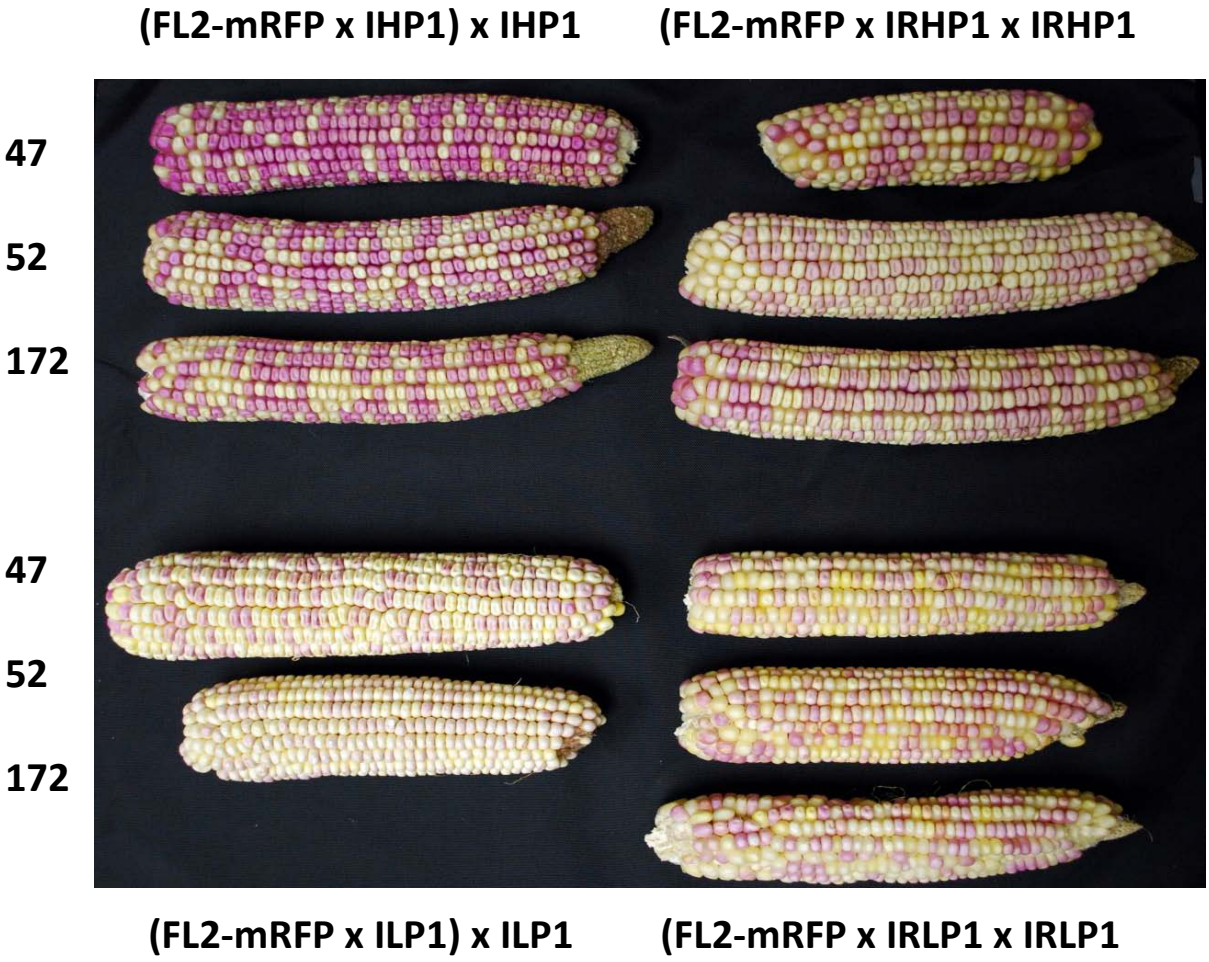


Figure 6. BC1 crosses between the Illinois Protein Strains and the three *FL2-mRFP* reporter lines (47, 52 and 172). Segregation of kernel color is apparent by the observed variation in white-to-yellow coloration of non-mRFP kernels within an ear.

[(*FL2-mRFP* x IHP1) x IHP1]BC5

[(*FL2-mRFP* x ILP1) x ILP1]BC5

[(*FL2-mRFP* x IRHP1) x IRHP1]BC5

[(*FL2-mRFP* x IRLP1) x IRLP1]BC4

[(*FL2-mRFP* x B73) x B73]BC6

Figure 7. BC5 crosses between the Illinois Protein Strains and the event 52 *FL2-mRFP* reporter line.

Expression of *FL2-mRFP* throughout Development

Because the *FL2-mRFP* reporter lines provide a non-destructive visual assay for zein gene expression, transgenic seeds from the crosses to the IPS and B73 were monitored for the onset and progression of red coloration during seed development. Zein accumulation is known to begin around 10 DAP, peak at approximately 16 DAP, and continue throughout grain fill. To document mRFP expression throughout development, BC3 kernels from reciprocal crosses of the Illinois Protein Strains and all three transgenic reporter lines were photographed at various time points from 8 to 24 DAP, where the same ear was used for the entire series of photographs (**Figure 8**). No differences were apparent due to transgenic event or direction of cross. For this reason, only ears from event 52 are shown using the Illinois Protein Strains as the female parents. From these photographs, *FL2-mRFP* expression can be visualized beginning at 12 DAP in IHP and IRLP, consistent with the known patterns of zein accumulation,

and 14 DAP in ILP, IRHP and B73. Expression increases throughout development in all genotypes. These results indicate normal patterns of zein accumulation in the endosperm for all genotypes.

Like BC1 ears, BC3 ears also illustrate *FL2-mRFP* expression corresponding with known levels of zein protein accumulation; IHP1 exhibits dark pink, ILP1 light pink, and IRHP1 and B73 medium pink kernels. However, IRLP1 kernels demonstrate much darker pink coloration than is expected given the low protein concentration of this strain. Since these ears are only at the BC3 stage of backcrossing, they contain an average of 93.75% of the recurrent parent genome and 6.25% of the transgenic genome, meaning they could still contain alleles controlling protein concentration from the transgenic parent genome.

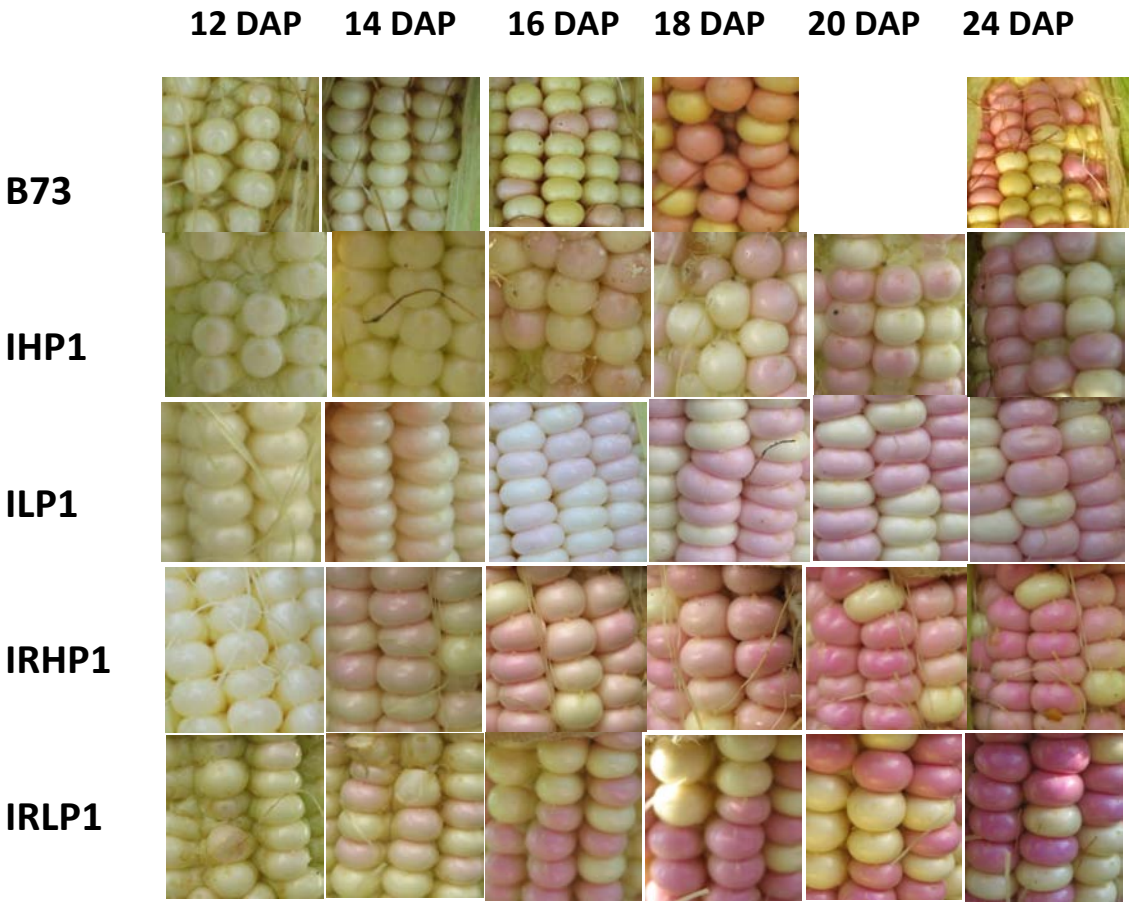


Figure 8. Photographs of BC3 kernels of the inbred-derived Illinois Protein Strains (event 52) and B73 (event 47) at 12, 14, 16, 18, 20 and 24 days after pollination (DAP).

Quantification of *FL2-mRFP* Expression

Adobe Photoshop was used to quantify the magenta channel in the photographs of the developing kernels (**Figure 8**) to provide estimates of *FL2-mRFP* expression. The Photoshop intensity scale ranges from zero to 256, where zero indicates no detection of magenta (white). Increasing magenta channel values signify an increasing amount of *FL2-mRFP* expression in the kernels. **Figure 9** shows the levels of magenta detected in all genotypes at three time points throughout development (12, 16 and 22 DAP) where a standardized area of 15 x 12 pixels was measured for four kernels per genotype. The means of the four replicates are graphed and the error bars indicate the standard errors. All genotypes exhibit increasing magenta channel intensity throughout development, consistent with visual observations. At all time points, magenta signal is strongest for IRLP1, followed by IHP1, IRHP1, B73 and ILP1, with the exception of B73 levels at 12 DAP, which are slightly higher than IRHP1. The magenta channel intensities detected by Photoshop corresponded with visual detection, and correlated with known protein concentrations of all genotypes, except for IRLP1. The measured magenta channel intensities from 22 DAP kernels was highly correlated (0.81) with known protein concentrations of mature kernels. IRLP1 seems to be an anomaly, where the kernels demonstrate stronger than expected pink coloration based on their protein concentration. Despite containing only half of the protein as IHP1, the detection of pink coloration is nearly as high as that measured for IHP1 at 22 DAP and exceeds that of IHP1 at 16 DAP. It is possible that this could be attributed to the early stage of backcrossing, or may reveal biological differences in zein regulation within IRLP1. Collectively, these results demonstrate that it is possible to quantify differences in the pink coloration of the photographed kernels by Adobe Photoshop and they generally follow known patterns of zein accumulation throughout development.

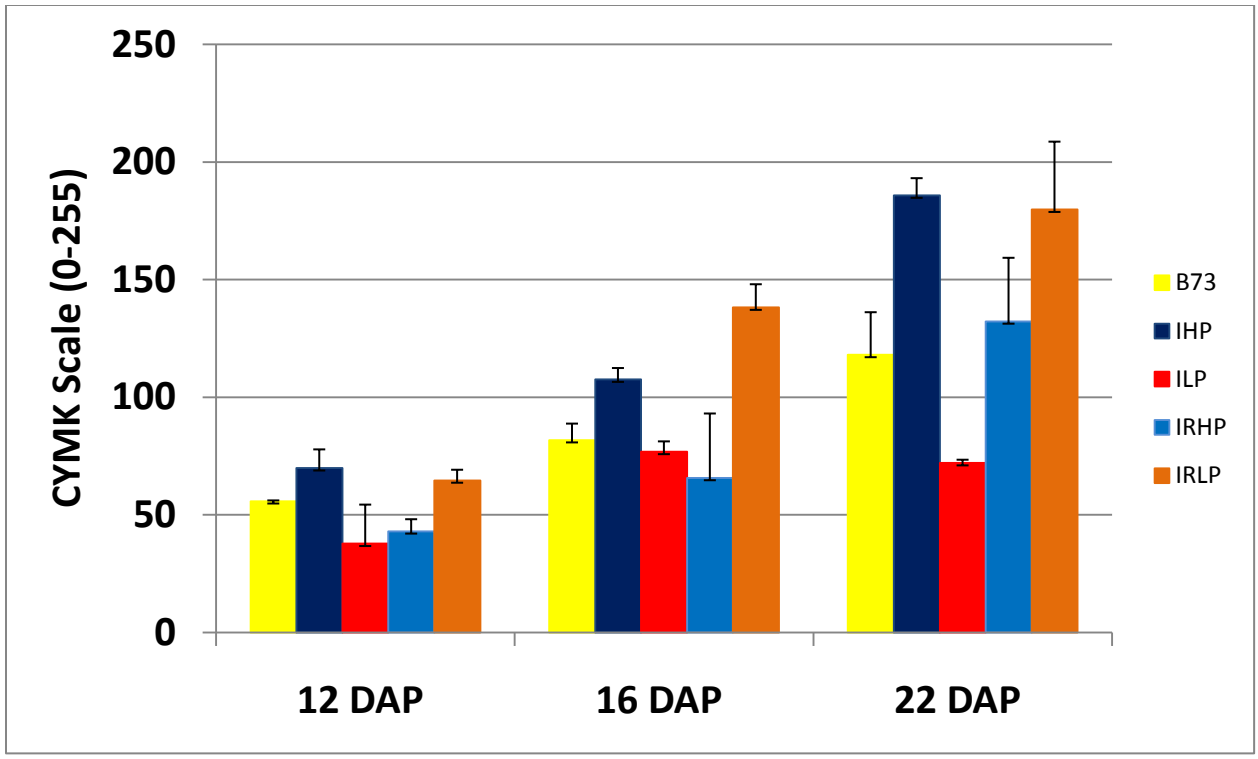


Figure 9. Adobe Photoshop quantification of the magenta channel intensity in photographs of developing BC3 kernels (Figure 5) to estimate *FL2-mRFP* expression. The scale ranges from 0-255, where 0 corresponds with no signal and 255 corresponds with saturation of the magenta channel.

To further corroborate mRFP pink coloration intensities with gene expression, zein accumulation and relative expression of zein and *FL2-mRFP* transgene expression were examined at 24 DAP in the pink seeds from BC3 ears in IHP and ILP backgrounds and are shown in **Figure 10**. As expected, zein accumulation was higher in kernels from the IHP background compared to ILP. Using qRT-PCR with primers that specifically amplify each α-zein subfamily, the Z19A, Z19D and Z22C RNA expression was much higher in IHP compared to ILP, but Z19B showed similar expression in the two backgrounds. Importantly, the relative expression of *FL2-mRFP* was also higher in the IHP background than ILP, with the ratio of IHP to ILP expression (3.1) being similar to that observed for Z22C genes (3.4), of which Fl2 is one of the more abundantly-expressed members.

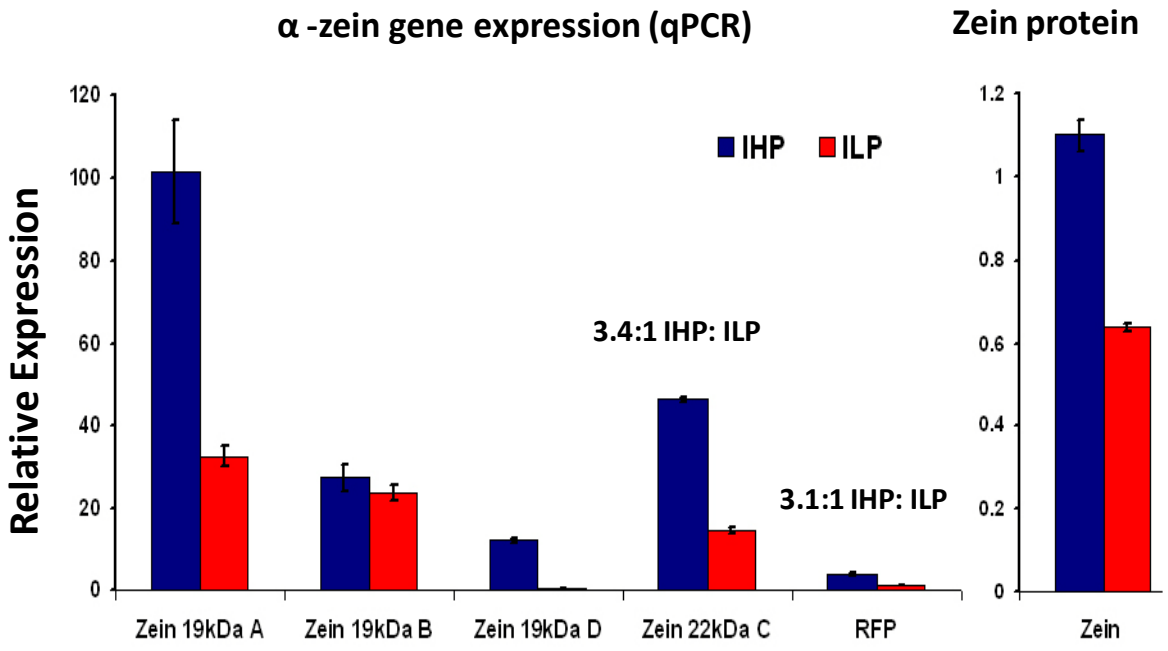


Figure 10. Relative expression of 19 kD (A, B, D), 22 kD (C) α-zein genes and *FL2-mRFP* in IHP1:*FL2-mRFP* and ILP1:*FL2-mRFP* BC3 kernels.

Regulation of *FL2-mRFP*

The *o2* recessive mutant was backcrossed repeatedly into IHP1 to create a near-isogenic line for *o2* in the IHP1 background (IHP1: *o2*). The *FL2-mRFP* in the IHP1 background (BC4 stage) was then backcrossed twice to IHP1:*o2*to generate ears segregating for both the *FL2-mRFP* transgene and *o2*. The resulting ear is shown in **Figure 11**. Kernels homozygous for the *o2* mutant allele (*o2/o2*) have a chalky, opaque phenotype compared to their wild-type counterparts (O2/?), which is characteristic of decreased zein accumulation in the endosperm. Approximately 50% of the kernels contain the *FL2-mRFP* transgene, as visualized by the pink coloration of the kernels. As expected, approximately 25% of the kernels are homozygous for the o2 mutation and contain the *FL2-mRFP* transgene (*o2/o2*; RFP), and these kernels appear opaque and light pink in color when compared to kernels containing only the *FL2-mRFP* transgene (O2/?; RFP). The reduction in pink coloration indicates a reduction in *FL2-mRFP*

expression and a corresponding decrease in zein protein accumulation by *o2*. These results suggest that the *FL2-mRFP* transgene is regulated by *Opaque2* in a similar fashion as the endogenous *Floury2* gene.

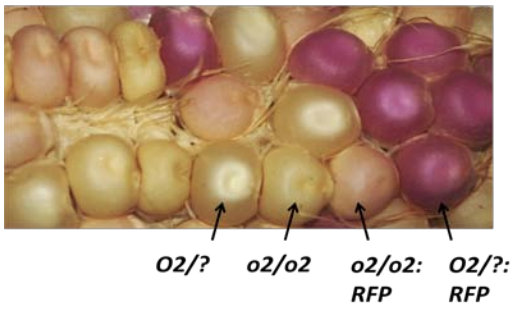


Figure 11. Photograph of an IHP1 ear segregating for the *o2* mutation and the *FL2-mRFP* transgene. The *o2* mutant phenotype is chalky and opaque in appearance (*o2/o2*) compared to wild-type kernels (O2/?). Kernels containing both the *o2* mutation and the *FL2-mRFP* transgene (*o2/o2*; RFP) illustrate reduced expression of *FL2-mRFP*, as detected by a significant decrease in pink coloration of the kernels compared to kernels containing only the *FL2-mRFP* transgene, which are much darker (O2/?; RFP).

FL2-mRFP to Study Protein Body Formation

Zein proteins assemble into protein bodies and accumulate in the subaleurone and starchy endosperm. It has been shown that IHP contains more RNA and protein body bound polysomes, as well as a higher efficiency in the ability of IHP polysomes to translate the RNA into protein (Miflin and Zoschkefeds, 1978). Therefore, one hypothesis for increased zein expression in IHP versus ILP is that there are simply more protein bodies present, or that they are larger in size. An alternative hypothesis is that a greater number of cell layers contain protein bodies. The presence of the *FL2-mRFP* transgene in the IPS allowed for comparisons of zein gene expression at the cellular level between IHP1 and ILP1. BC5 kernels were available for analysis. At this stage of backcrossing, kernels contained an average of 98.43% of the recurrent parent genome. Kernels were collected at 8, 12, 16, 20 and 24 DAP. However, only comparisons of 24 DAP kernels were possible due to the extremely low amounts of fluorescence detected in ILP1 at early time points. While it was possible to increase the power of the laser to excite *FL2-mRFP* at a higher level in ILP1, doing this saturated the detection of fluorescence in IHP1 and complicated comparisons between the two genotypes. Therefore, the same power was used for exciting *FL2-mRFP* in both genotypes.

Images of the kernels were taken using the Zeiss confocal laser scanning microscope (LSM 710). One feature of this microscope is the ability to combine aspects of light and fluorescence microscopy. The images below take advantage of this feature, where light microscopy allows for the visualization of the different cell layers, including the pericarp, the aleurone and the subaleurone, and fluorescence microscopy allows for the visualization of *FL2-mRFP* expression within the different cell layers. Utilizing a 40x objective, detailed images were taken of 24 DAP IHP1 (**Figure 12a**) and ILP1 kernels (**Figure 12b**). From these images and others like them, it is possible to quantify the number of cellular layers expressing the *FL2-mRFP* transgene. No expression is evident in the pericarp, aleurone or embryo, consistent with known zein accumulation exclusively in endosperm tissue. In both IHP1 and ILP1, *FL2-mRFP* expression is the strongest in subaleurone cells, with a progressive reduction in expression into the starchy endosperm. From these images it was determined that *FL2-mRFP* expression extends much further into the kernel in IHP1 compared to ILP1. Greater than twenty cell layers of the subaleurone and starchy endosperm exhibited *FL2-mRFP* expression in IHP1, while only approximately five layers exhibited expression in ILP1.

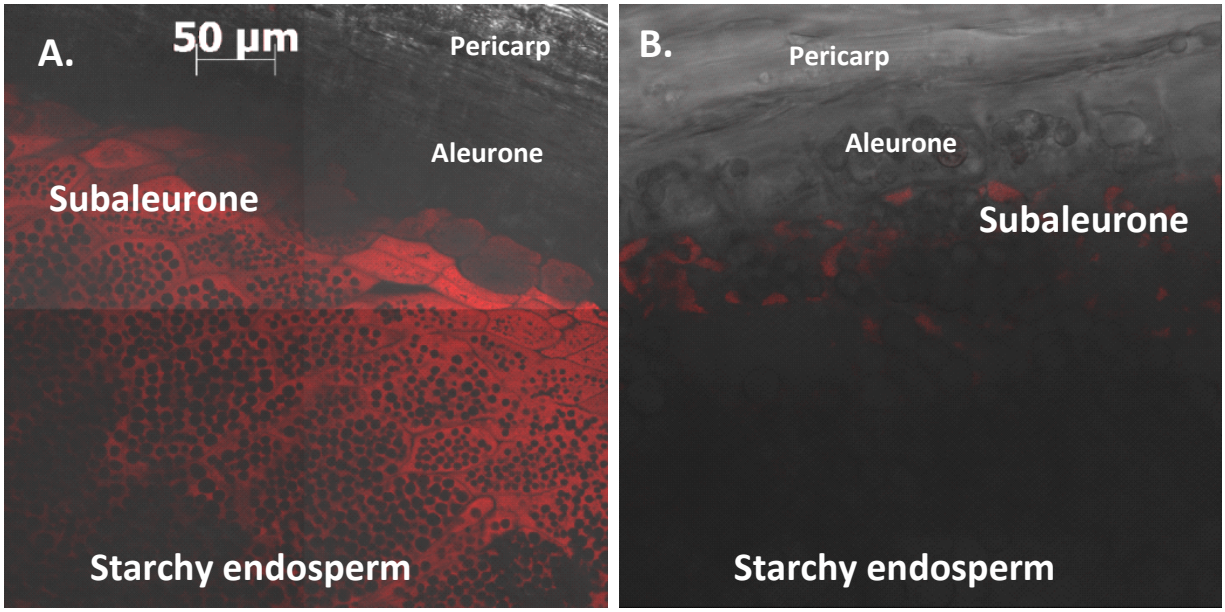
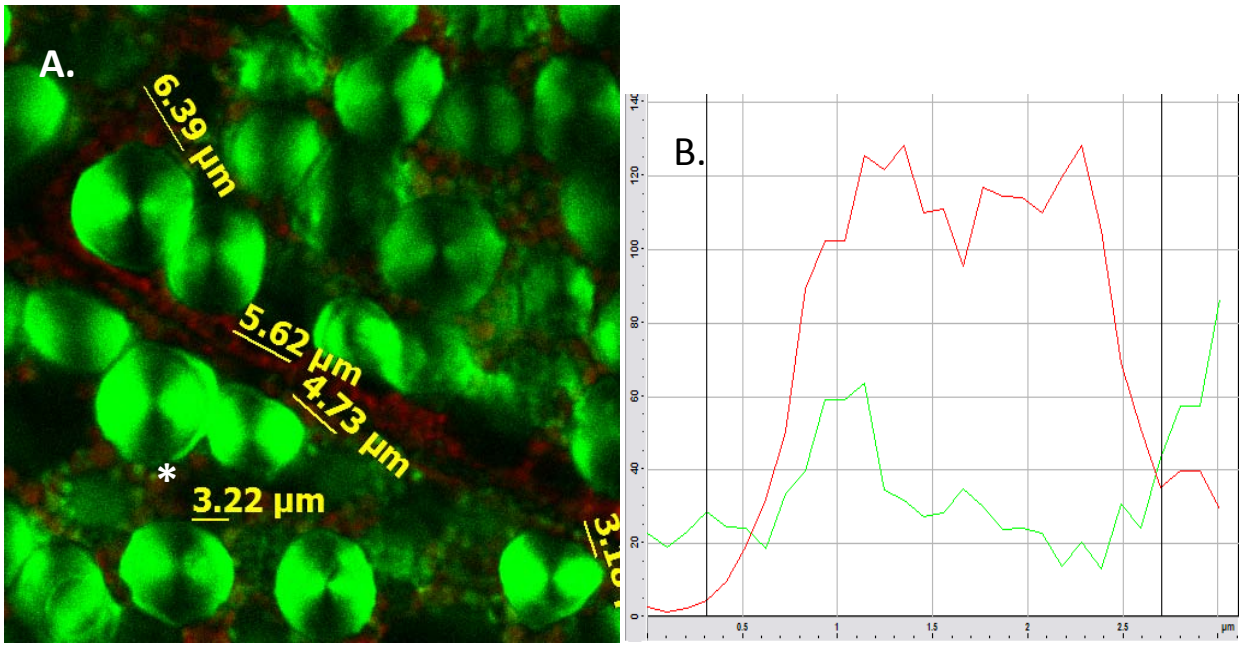


Figure 12. Detection of *FL2-mRFP* fluorescence within cells in IHP1 (A.) and ILP1 (B.) 24 DAP kernels by light and fluorescence confocal microscopy.

It was also possible to view individual protein bodies in IHP1 and ILP1 kernels, as illustrated in **Figure 13** below. Individual protein bodies could be identified by their expression patterns, where each body exhibited a characteristic peak in expression ranging in diameter from 1-3 microns, as measured using the line/intensity profile tool in Axiovision. Based on diameter, and the spherical nature of the bodies, the volumes were calculated. The average volume of protein bodies was 5.19 μ^3 in IHP1 and 4.5 μ^3 in ILP1, indicating no significant difference between HP1 and ILP1 (**Fig. 13C**). On the other hand, IHP1 contained a much greater number of bodies than ILP1. The software utilized to analyze the images also allowed for the detection of starch granules by their characteristic patterns of concentricity. Here, they are depicted in green and could be measured in the same way the protein bodies were measured. However, similar to protein bodies, the average size of starch granules did not vary between IHP1 (842 μ^3) and ILP1 (906 μ^3).

Crosses of the *FL2-mRFP* Reporter Lines to the IPS to Study the Maternal Effect on Grain Protein Concentration

To test for the source of the maternal effect, crosses between the IPS and the *FL2-mRFP* transgenic lines were made. To investigate the possibility of maternal genomic imprinting, reciprocal crosses were made between each of B73, IHP1 and ILP1 inbreds and the *FL2-mRFP* reporter lines. At this point, the *FL2-mRFP* transgene had been backcrossed five generations to B73 and four generations to IHP1 and ILP1 to create BC6 (B73) and BC5 (IHP1 and ILP1) ears **(Figure 2)**. If maternal genomic imprinting was the cause of the maternal effect in the Illinois Protein Strains, *FL2-mRFP* expression would only be expected when transmitted through the female parent. All ears resulting from these crosses result in *FL2-mRFP* expression, as judged by the red coloration of the kernels (**Figure 2a**). The color intensity is consistent with the zein protein concentrations in each of these lines, where IHP1 accumulates the most zein protein and produces the darkest red kernels, ILP1 accumulates the least zein protein and produces the lightest pink kernels, and B73 accumulates intermediate zein protein and produces intermediate pink



C.	Cell layers	PB d	PB vol	std dev.	SG diam	SG vol	std dev.
IHP	>20	2.10	5.19	1.86	11.62	842.54	256.30
ILP	5	2.04	4.50	1.20	11.52	906.12	516.88

Figure 13. Detection of individual protein bodies and starch granules utilizing the 40x objective. A. Visualization of protein bodies by detection of *FL2-mRFP* expression (red) and starch granules by detection of concentric patterns (green) in 24 DAP IHP1 kernels. The diagonal line (upper left to lower right) formed by red protein bodies indicates the border of two starchy endosperm cells. B. Expression level of *FL2-mRFP* (red line) and starch granules (green line) across the 3.22 µ increment indicated by the asterisk in A. The diameter of a protein body can be measured based on the width of the red peak. This protein body is approximately 2.5 µ in diameter. Similar analysis was done to measure the diameters of starch granules. C. Summary of number of cells expressing *FL2-mRFP*, and protein body and starch granule size between IHP1 and ILP1 24 DAP kernels.

kernels. The observed differences in *FL2-mRFP* expression are expected given the advanced stage of backcrossing of the transgene to the strains. However, the ears created from the reciprocal crosses, where the transgene was inherited through the male parent, also exhibit *FL2-mRFP* expression, where pink coloration also follows protein concentration (**Figure 2b**). These results indicate that neither maternal nor paternal imprinting is the cause of the maternal effect in the IPS.

The second possible cause of the maternal effect is endosperm dosage, which can also be investigated using the results of the previously described crosses between the *FL2-mRFP* reporter lines

and B73, IHP1 and ILP1 (**Figure 14**). If dosage effect was responsible for the maternal effect, we would expect to see up to two times stronger *FL2-mRFP* expression in ears created using the *FL2-mRFP* genotype as the female parent because kernels from those ears receive two copies of the transgene. However ears created using *FL2-mRFP* as the female parent (**Figure 14a**) show approximately equivalent expression as ears created using the transgenic lines as the male parent (**Figure 14b**). However, B73 ears do appear to exhibit stronger *FL2-mRFP* expression when the transgenic line is used as the female, which may provide evidence for a small dosage effect in B73. On the other hand, dosage effect is not apparent in IHP1 or ILP1 ears, which could indicate the absence of a dosage effect or that the magnitude of its effect is relatively small compared to other more significant mechanisms. Overall, these results indicate that dosage effect is not a major cause of the maternal effect in IHP1 or ILP1, but may influence B73. This suggests the possibility for variation in the magnitude of the maternal effect depending on genotype.

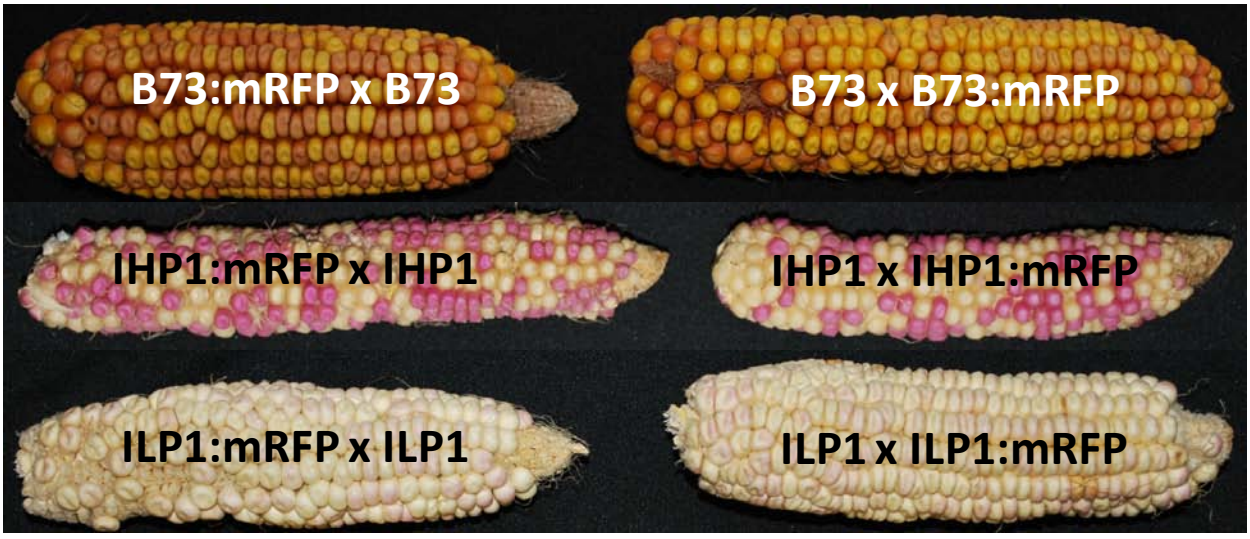


Figure 14. Photographs of BC5 ears (IHP, ILP) and BC6 ears (B73) obtained from reciprocal crosses between *FL2-mRFP* transgenic plants and the IHP1, ILP1 or B73 backgrounds. The ears on the left (A.) were created using the *FL2-mRFP* lines as the female parent. Each of these ears contains two copies of the transgene. The ears on the right (B.) were created using the *FL2-mRFP* lines as the male parent. Each of these ears contains only one copy of the transgene.

The third possible cause of the maternal effect is the nitrogen supply from vegetative source tissues, which is dependent upon the genotype and determines the composition of assimilates entering the seed. In order to investigate this possibility, crosses were made holding the maternal genotype constant as a control and varying the genotype of the male parent, and vice versa (**Figure 15**). Since B73 is not under selection for protein concentration, B73:*FL2-mRFP* was used as the control genotype. At this point, the *FL2-mRFP* transgene had been backcrossed five generations to B73, so the plants used in this experiment are expected to exhibit only minimal genetic variation for their nitrogen status. The first set of crosses was made using B73:*FL2-mRFP* as the female and the B73, ILP1 or IHP1 inbred lines as the males. All three ears created by these crosses, regardless of paternal genotype, exhibit equivalent *FL2-mRFP* expression that is consistent with the accumulation of zein protein in B73 ears (**Figure 15a**). The ears created from reciprocal crosses, however, demonstrate large differences in *FL2-mRFP* expression that is consistent with the zein protein accumulation of the maternal genotype. IHP1 plants produce the darkest red kernels, ILP1 plants the lightest, and B73 intermediate, but significantly lighter than IHP1 kernels (**Figure 15b**). The results of these crosses indicate that the nitrogen status of the plant may be largely responsible for the maternal effect on grain and zein protein accumulation in B73, IHP1 and ILP1.

(B73: *FL2-mRFP*)BC5 Crosses to the IPSRI Population

Dudley et al (2007)previously generated a population of 500 families derived from the cross of IHP and ILP plants at cycle 70, followed by seven generations of random mating. This population has been used in genetic mapping experiment to identify QTL controlling grain composition (Dudley et al., 2007). The Moose laboratory continued to inbreed the 500 families in this population to create a set of recombinant inbred lines that permit replication of defined genotypes and will be useful in a combined linkage and association mapping approach. This population of Illinois Protein Strain Recombinant Inbreds (IPSRIs) exhibits grain protein concentrations ranging from 5-24%. Here, we crossed the B73 transgenic reporter line (event 52, BC5) to a subset of 138 individuals whose protein concentrations

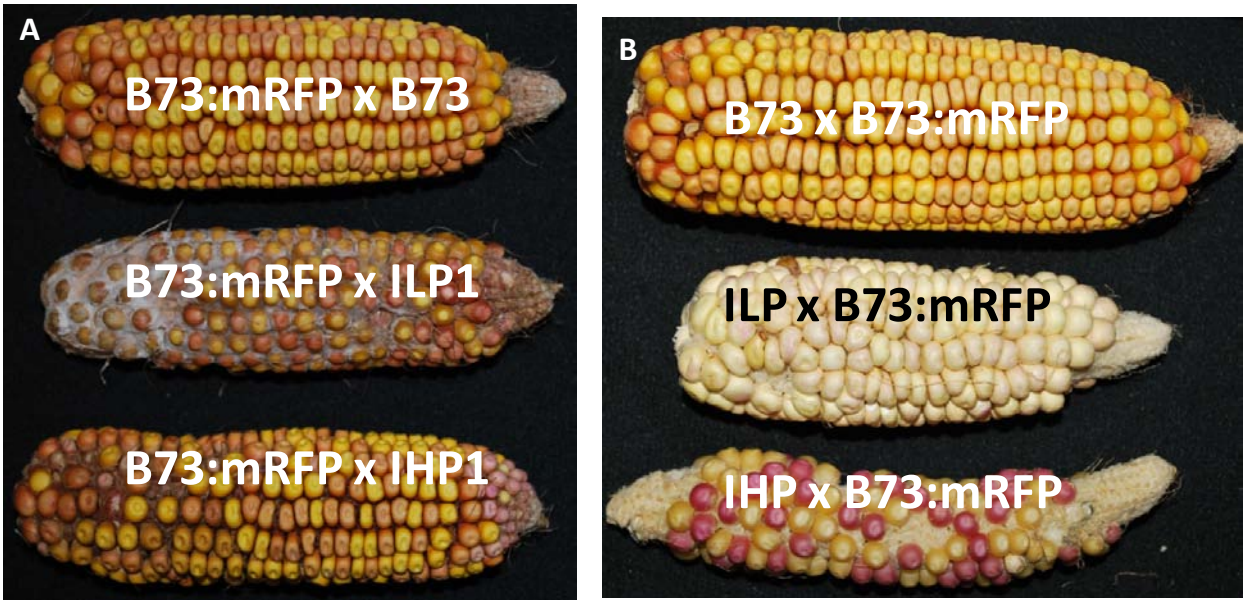


Figure 15. A. Photographs of ears created by crossing B73, ILP1 and IHP1 to B73:mRFP. B. Photographs of ears created by crossing B73:mRFP to B73, ILP1 and IHP1.

span the dynamic range of the full 500 lines, and also most fully represent the genetic variation in the original 500 lines based on genotyping with a set of 500 SNP markers (Dudley et al., 2007). Frequencyhistograms of the original 500 lines and the subset of 138 lines are plotted according to protein concentration in **Figure 16**. Future uses for the *FL2-mRFP* reporter lines would most likely involve detection of small differences in protein concentration in elite inbred lines, whose protein concentrations range from 8-12%. B73 contains approximately 7% protein, which is closer to the 8-12% range of the elite inbred lines than any of the other reporter lines. For this reason, the B73 reporter line was used in the crosses to the IPSRIs. The resulting ears demonstrate a wide range of pink coloration (**Figure 17**) and are awaiting quantification.

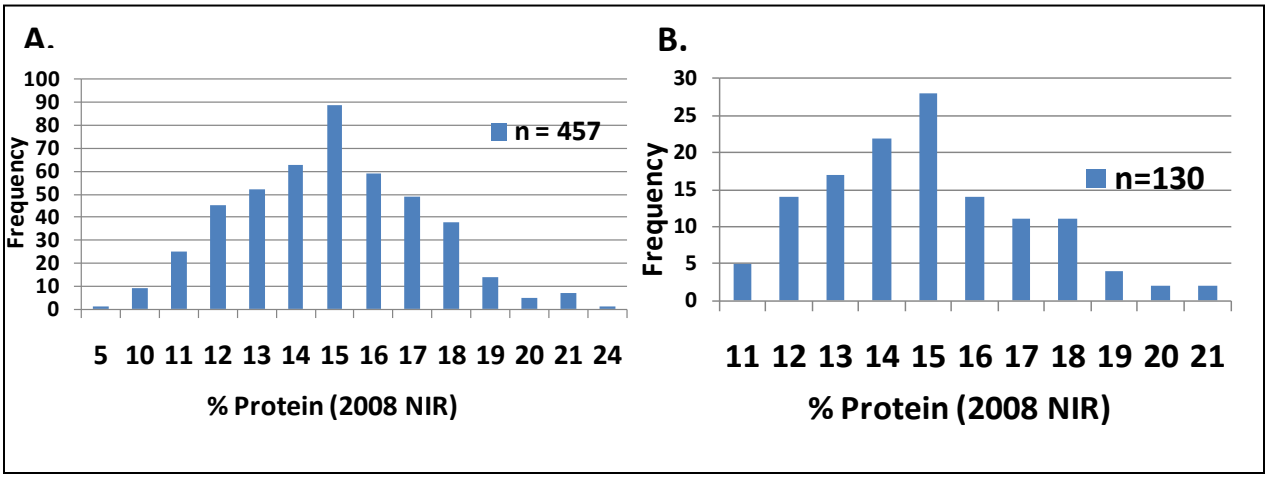


Figure 16. Frequency histograms of the Illinois Protein Strain Recombinant Inbred (IPSRI) population protein concentr6tions as measured by NIR in 2008. A. Histogram of 457 individuals. B. Histogram of the sub-population used in this experiment.

Figure 17. Photograph of three ears created by crossing B73: *FL2-mRFP* (BC5) pollen onto three IPSRIs of varying protein concentrations, as indicated by the variation in *FL2-mRFP* expression in these ears.

DISCUSSION

Prior research has established that selection for grain protein concentration in the Illinois Protein Strains has altered α-zein accumulation, but characterization of regulatory factors underlying these changes are not well understood. Molecular studies of zein genes are typically complicated due to the high number and sequence similarity among this large multi-gene families. Fluorescent protein reporter lines may overcome some of these issues because they allow for tracking of individual zein gene expression. Recent advances in fluorescent protein technology have generated improved variants with enhanced photo stability, faster maturation, and which cover a broader spectral range. The additional ability of the DsRed fluorescent protein to be visualized under white light is a feature that eliminates the necessary destruction of the tissue under study. For these reasons, transgenic reporter lines are becoming more widespread in plant biotechnology for investigating protein localization and gene expression. The studies in this thesis evaluate the utility of the *FL2-mRFP* reporter lines in characterizing α-zein gene regulation when crossed to the Illinois Protein Strains.

FL2-mRFP expression in the Illinois Protein Strains can be visualized under white light, as detected by the intensity of pink coloration of the kernels, allowing for the quick non-destructive monitoring of α-zein protein accumulation throughout development. Quantification of the intensity of pink coloration of photographed kernels is possible using Adobe Photoshop by measuring the magenta channel. Photoshop measurements are in accordance with visual estimates, where both correlate with known protein concentrations in all of the genotypes, except for IRLP1, and increase throughout development, consistent with known patterns of zein accumulation. It is not clear as to why IRLP1 exhibits such intense pink coloration. At this point, IRLP1 had only undergone three generations of backcrossing, so it is feasible that the BC3 genome still contained an average of 6.25% of the transgenic donor parent (Hi-II x B73) genome and also possibly donor parent alleles for protein concentration. Another possibility is that the regulatory changes in α-zein gene expression that have occurred in IRLP

affect the *FL2-mRFP* transgene activity (e.g. zein turnover) in ways that are not directly proportional to the final protein accumulation.

Although Photoshop measurements did show a strong correlation with protein concentration in the Illinois Protein Strains, several sources of variation need to be eliminated to increase the robustness of this technology. First, genetic variation could be eliminated by further backcrossing, and photographs could be taken of BC6 or BC7 ears to control for this. However, perhaps the largest source of variation is the environmental variation under which the photographs were taken. Ideally, ears should be photographed under standardized conditions. Future work will involve the building of an enclosed structure for photography where individual kernels or ears may be photographed at a standard distance and using a filter that only collects a narrow range of wavelengths corresponding to *FL2-mRFP* emission spectra. It would also eliminate differences in the emission spectra due to variation in kernel color across the strains, which can be observed in the kernels not expressing *FL2-mRFP* and which causes differences in the white-yellow coloration of the kernels. Additionally, it will be necessary to create a standard curve containing genotypes that exhibit a fuller range of protein concentrations. Doing this would accomplish two goals: first, it would provide a better idea of how closely the detected levels of pink coloration correlate with protein concentrations, and thus establish whether protein concentrations can be estimated based solely on the pink coloration of the kernels. Secondly, it would establish the sensitivity of the Adobe Photoshop software. Here, we established that Photoshop can detect large differences in coloration associated with the large differences in protein concentration of the Illinois Protein Strains, but it is not known if it can detect small differences in pink coloration and, thus, protein concentration. The IPSRI population will be useful for this purpose, as it provides a dynamic range of protein ranging from 11-21%.

The cluster arrangement of the 22 kD α-zein genes has been hypothesized as one underlying cause of their coordinated regulation. Comparisons of *FL2-mRFP* expression across transgenic events

may provide insight with regards to this hypothesis. Since the *Agrobacterium* method of transformation inserts randomly in the genome, the probability of any one of the events landing in the zein gene cluster is low, let alone all three events landing there. However, significant differences in *FL2-mRFP* expression, as detected by the pink coloration of the kernels, were not observed among transgenic events for any of the genotypes. This data suggests that the cluster arrangement of the 22 kD α-zein genes is not important for their coordinated regulation. Furthermore, it appears that the mechanisms controlling *FL2-mRFP* gene expression are the same for all three transgenic events. These results suggest that a trans-acting factor may be regulating zein gene expression. While it is evident from our data that *Opaque2* is one trans-acting regulatory mechanism controlling *FL2-mRFP* gene expression in a similar manner as the endogenous *Floury2* gene*,* the *o2* null mutant did not completely reduce *FL2-mRFP* expression. This indicates the presence of other regulators. Additionally, the *O2* binding site is only present in the promoters of the 22 kD α-zein genes, not the 19 k kD α-zeins, so the mechanism that is coordinately regulating all α-zeins cannot be *O2*, but some other putative transacting factor.

Crosses between the reporter lines and protein strains have allowed for the study of the maternal effect on grain protein concentration in the Illinois Protein Strains. Three possible mechanisms underlying the maternal effect have been proposed: maternal imprinting of zein gene expression, endosperm dosage effects, and nutrient supply from vegetative source tissues (Moose et al., 2004). If maternal imprinting was the source of the maternal effect, *FL2-mRFP* expression would only be observed when transmitted through the female parent. However, ears produced from both directions of crosses (**Fig. 14**) demonstrate *FL2-mRFP* expression, regardless of whether it is inherited through the female or male. This data suggests that maternal imprinting does not cause the maternal effect. Analysis of the *level* of expression in these ears can also provide information regarding dosage effect. Due to the triploid nature of the endosperm, the *FL2-mRFP* transgene may vary in copy number from zero to three, depending on whether it is transmitted through the female or male, and if it is

heterozygous or homozygous. Reciprocal crosses of plants heterozygous for the transgene to the Illinois Protein Strains generated ears expected to segregate either for zero and one *FL2-mRFP* allele (if transmitted through the male) or zero and two alleles (if transmitted through the female). If dosage effect were the main cause of the maternal effect, ears that have inherited the transgene through the female would be expected to illustrate stronger *FL2-mRFP* expression than ears inheriting the transgene through the male because they contain an extra copy of the transgene. However, no differences were observed in the *level* of *FL2-mRFP* expression in either of the IHP1 or ILP1 backgrounds, which suggests that dosage effect does not significantly contribute to the maternal effect in these genotypes. On the other hand, B73 did illustrate stronger *FL2-mRFP* expression when the transgene was inherited through the female, so dosage effect might contribute to the maternal effect in B73.

The lack of support for either maternal imprinting or endosperm dosage in IHP1 and ILP1 leaves the nutrient supply from the source tissues as the most likely source of maternal control of zein gene expression and protein accumulation. Prior studies have shown that the nutrient supply is dependent upon genotype, where IHP plants demonstrate elevated N uptake, N assimilation by seedling leaves, and N remobilization from source to seed sink tissues compared to ILP plants (Below et al., 2004). Ears produced from reciprocal crosses between IHP1, ILP1 and B73 with a control genotype (B73:mRFP) illustrate dramatic differences in *FL2-mRFP* expression when the inbreds are used as female parents, but not when the control genotype is used as the female, where IHP1 illustrates the strongest, B73 intermediate, and ILP1 the least *FL2-mRFP* expression (**Fig. 15**). These observations are consistent with what is known about the relative N content of the vegetative tissues from these genotypes. Additional observations supporting the nutrient supply hypothesis include the consistent *FL2-mRFP* expression observed on all kernels within an ear, regardless of genotype and stage of backcrossing (**Fig. 7, 8, 14, 15**). This observation suggests that the phenotype is controlled by a mechanism that is acting on the entire ear, even when the kernels on an ear are segregating, such as in the BC1 population (**Fig. 7**).

Additional experiments will be required to confirm the nutrient supply hypothesis. Two possible experiments are to test the effects of different levels of N fertilizer on *FL2-mRFP* expression on field- or greenhouse- grown plants, or the effects of N supply on *FL2-mRFP* expression in kernel culture experiments. However, in order to control for variation in zein accumulation due to genetic segregation, it is important to use fully backcrossed genotypes (BC6 or BC7 plants) in these experiments. It would also be interesting to test for the source of the maternal effect in IRHP1 and IRLP1 to see if the nutritional status of the plant is as important in these genotypes as it is in IHP1 and ILP1. The *FL2-mRFP* reporter lines will also be useful for studies where it is desirable to track protein concentration quickly and non-destructively. Relative *FL2-mRFP* expression levels are already being used as phenotypes in quantitative trait mapping studies to identify genes that may regulate α-zein gene expression and protein accumulation. The *FL2-mRFP* reporter lines could also be used in mutagenesis experiments, where a reduction in expression would indicate the presence of a mutation disrupting the function of gene(s) regulating protein concentration.

MATERIALS AND METHODS

FLOURY2-mRFP TRANSGENIC LINES

The FLOURY2-mRFP transgenic reporter lines used in this research are one of a series of transgenic reporter maize lines that were created as part of a joint project conducted by Dave Jackson's laboratory at Cold Spring Harbor, Anne Sylvester's laboratory at the University of Wyoming, and the Plant Transformation Center at Iowa State University in order to characterize protein localization and expression in maize. The reporter lines are created by fusing fluorescent reporter genes to a set of gene sequences whose encoded proteins represent comprehensive coverage of cellular compartments (see http://maiaze.tigr.org/cellgenomics/index.shtml). The Floury2-mRFP construct consists of a modified red fluorescent protein (mRFP) fused to the C-terminus of the *FLOURY2* 22 kD α-zein protein, , and serves as a marker for protein bodies in maize endosperm cells. To preserve proper tissue and temporal regulation of the FLOURY2-mRFP fusion protein, the construct is driven by native flanking regulatory elements, including approximately 2000 bp of the *Floury2* gene promoter and 1000 bp of 3' sequence. Complete methods for generation and analysis of the fluorescent protein-tagged maize lines can be found in Mohanty et al. (2009).

The *Agrobacterium* method was used to transform embryogenic callus of the Hi-II genotype with a T-DNA that also contained the *bar* gene under the control of the CaMV 35S promoter as a selectable marker. Approximately 20 seeds from three transgenic events, named 47, 52 and 172, expressing the *FL2-mRFP* construct were obtained from Dr. Dave Jackson, Cold Spring Harbor.

Introgression of the mRFP transgene into the inbred-derived Illinois Protein Strains was begun by Salas (2008) by crossing the mRFP reporter lines to the inbred lines IHP1, ILP1, IRHP1, IRLP1 and B73. The hybrids generated from these crosses were backcrossed to the Illinois Protein Strains and B73, and BC1 seed was available at the start of this project. Since the presence of the transgene can be identified easily and non-destructively by the pink coloration of the kernels, only pink BC1 kernels were planted for

further backcrossing. The Illinois Protein Strains and B73 were planted at a 10-day delay to account for hybrid vigor of the BC1s. Kernels were planted in the greenhouse and grown to maturity under a daily light/temperature regime of 16 hours at 28 degrees C and eight hours at 22 degrees C. Pollinations were conducted in both directions at each backcross until the BC4, but only seeds where the inbred line was the female parent were planted for the next generation of backcrossing. Subsequent backcrosses were created by alternating plantings in the greenhouse (winter 2008, 2009) under the conditions described above and the 2009 and 2010 summer nurseries. Additionally, a null mutant allele of *opaque2* (*o2-R*) was backcrossed into IHP1 to create the IHP1: *o2* near-isogenic line. IHP:*FL2-mRFP* plants from the BC4 generation were then backcrossed to the IHP1:*o2* genotype to create a BC5 ear segregating for *o2* and *FL2-mRFP* in the IHP1 background.

To ensure stable behavior of at least one of the three reporter lines, all three of the transgenic events (47, 52 and 172) were backcrossed until BC4 in all strains. Analysis of the BC4 IPS:*FL2-mRFP* seed, however, revealed no differences in *FL2-mRFP* expression patterns due to transgenic event, so only event 52 was backcrossed further to the Illinois Protein Strains. All three transgenic events were fully backcrossed to B73. BC5 kernels for the Illinois Protein Strains will be planted in the greenhouse in the 2010 nursery to generate BC6 ears.

INITIAL IMAGING OF DEVELOPING mRFP KERNELS AND mRFP QUANTIFICATION

In order to document expression patterns of the *FL2-mRFP* reporter, photographs were taken of the developing kernels at 12, 14, 16, 18, 20 and 22 days after pollination. At this point, the *FL2-mRFP* transgene had been backcrossed three generations into the Illinois Protein Strains and four generations into B73. Plants from the 2009 summer nursery were chosen based on shared pollination dates among the genotypes and the same ears were used in the photographs throughout development. Adobe Photoshop and ImageJ were used t o measure the magenta (CYMK mode) or red (in RGB mode)

channels, respectively, in photographs of 12, 16 and 22 DAP developing kernels. A standardized area of 15 x 12 pixels was measured in each kernel and four kernels were measured per genotype. Means of the four replicates and standard errors were calculated. Since measurements were equivalent utilizing both software programs and Photoshop required fewer steps for quantification, it was chosen as the preferred software.

MICROSCOPY

To document *FL2-mRFP* expression at the cellular level, immature pink kernels were collected for microscopy. Plants from the 2010 summer nursery were chosen based on shared pollination dates among the genotypes. The same ear was used throughout development when possible; otherwise, ears with the same pollination dates were chosen. At least four kernels were extracted randomly from the ears at 8, 12, 16, 20 and 24 DAP. 8 DAP kernels served as negative controls, as zein genes are not expressed at 8 DAP. Kernels were sliced longitudinally with a razor blade so as to divide the embryo in half, suspended in O.C.T., and frozen on dry ice. Samples were stored in a light-sensitive container in the -20 degree freezer until needed for microscopy.

Images of 24 DAP kernels were taken using the Zeiss confocal laser scanning microscope (LSM 710) using a 2-photon laser. With this microscope it is possible to image a thin optical slice out of a thick specimen, a method known as optical sectioning. This feature was necessary due to the shattering of the frozen tissue when sliced with a cryotome, which probably resulted from the high protein concentrations of these samples, making them brittle. Therefore, the frozen half-kernel samples were thawed and imaged in their entirety.

A cryotome was used to slice the frozen kernel samples, which were then thawed and washed with a 1% Phosphate Buffer Solution. 2-Photon Confocal Laser Scanning Microscopy was performed using the Zeiss confocal laser scanning microscope model 710 (LSM 710) located in the Institute for

Genomic Biology Core Facilities Microscopy Facility, U of I, Urbana, IL under the direction of Mayandi Sivaguru. Settings were applied to measure the excitation and emission spectra of mRFP according to the goals of this research. Because the mRFP has an excitation of 584 nm and an emission of 607 nm, a filter was used to collect emission spectra in the 400-690 nm range. Laser power was set at 1% for all samples. Axiovision4 software (Carl Zeiss) was used to analyze the images for protein body and starch granule expression patterns. Approximately 10-15 protein bodies and starch granules were measured per genotype.

ZEIN EXTRACTION

Approximately 20 white seeds and 20 pink seeds of the Illinois Protein Strain: mRFP lines were ground to a fine powder. 100 mg of powder was suspended in a borate solution containing 2% of β-mercaptoethanol with constant shaking at 37°C overnight for total protein extraction, according to Larkin's Lab Protocols (http://ag.arizona.edu/research/larkinslab/). After centrifugation for 15 min at 14,000 rpm, 300 μL of the supernatant was transferred to new 1.7 mL tubes and the remaining supernatant was stored as total protein extract. To solubilize the zein protein, 700 μL of 70% ethanol was added to 300 μL of supernatant and incubated at 37° for 2 hr. After centrifuging, 100 μL of the supernatant containing the zeins was transferred to new 1.7 mL tubes. The remaining supernatant was poured off and saved as a source of zein solubilized in ethanol. The pellet containing the non-zein proteins was also saved, dried and resuspended in a borate/BME solution. The 100 μL of zein suspended in ethanol was centrifuged in a SpinVac overnight to evaporate the ethanol and the leftover zein pellet was resuspended in 100 μL distilled water.

BCA ASSAY

Total protein concentrations were quantified according to the Pierce® Microplate BCA Protein Assay Kit- Reducing Agent Compatible using a Hitachi U-2000 Spectrophotometer. A standard curve was created by plotting the absorbance values of the standards versus their concentrations (ug/mL). An R^2 value of 0.98 was achieved. Protein concentration values for unknown protein samples were determined by plotting their absorbance values on the standard curve.

SDS PAGE

SDS-polyacrylamide gels (15%, w/v) were prepared according to the NEXTGEL™ Electrophoresis System by Amresco. Total protein, non-zein protein and zein protein samples were diluted according to the NEXTGEL™ protocol with sample loading buffer and heat shocked for 5 min at 95°C. Gels were run at room temperature at a constant current of 100 mA for 2D electrophoretic separation of the denatured proteins. Gels were stained with Coomassie Brilliant Blue straining solution (BioRad) overnight, and then destained for 2-4 hrs.

RNA ISOLATION and CLEAN-UP

Immature pink kernels were collected, flash frozen with liquid N, and stored in the -80° C freezer until isolation. Frozen samples were ground in liquid N using a mortar and pestle. Additional grinding was done using a genogrinder at 1500 rpm for 1m. Approximately 0.1 g of frozen tissue was used for isolation. Since RNA isolation is generally low for seed tissue, Trizol LS is used instead of Trizol. Also, modifications have been made to the Trizol RNA Isolation Protocol. 750 μL Trizol LS was added to each sample and vortexed until homogenous. Incubated at 37° C for 10 min. Centrifuged for 5 min at 14,000 rpm. Transferred 800 μL of the supernatant to new 1.7mL microcentrifuge tubes. Here, 100 μL 5M NaCl was added to digest starch, which is abundant in seeds. Added 300 μL chloroform. Mixed by inverting

for 5 min. Centrifuged for 20 min at 14,000 rpm (using centrifuge in refrigerated room). 300 μL of the uppermost supernatant was transferred to new 1.7mL microcentrifuge tubes, to which 300 μL of 95% isopropanol was added, and mixed by inverting. Centrifuged for 10 min at 14,000 rpm. Poured out isopropanol and washed 3 times with 75% ethanol. After pouring off the ethanol, air dried the RNA pellets by blotting and inverting on paper. Re-suspended in 100 μL distilled water.

10 μg of the RNA samples were DNase treated. The total reaction volume was 50 μL. 5 μL each of DNase I and DNase I Buffer (NEB) were used. Samples were incubated at 37° C for 15 min. Following DNase digestion, the RNA samples were cleaned using the RNeasy RNA Clean-up commercial kit following manufacturer's instructions. Approximately 5 μg of RNA remained after DNase digestion and clean-up. Reverse transcription was performed using 1 μg of RNA per reaction. The total reaction volume was adjusted from 20 μL to 25 μL.

qRT-PCR

Primers were designed to anneal to common sequences of each subfamily of the 19 and 22 kD α-zeins, and the *FL2-mRFP* transgene: 19A Subfamily F:GTAGAACGGCAACAACTG; R: CAACAACAACAACTACCATTC; 19B Subfamily F:GCTGTGTCAAGAAGGAAG; R:CCATTCAACCAACTGTCTAC; 19D Subfamily F:GCAGATGGATGGCAGAAG; R:CACACAACAACTACAACAACAG; 22C Subfamily F:ACCAATGGGTTTGCCACTAC; R:GGGCAATGCACCTACATACC; mRFP F:GGCCCCAGCGGCCGCAGCAGCACCAGC; R:GGCCGGCCTGGAGGTGGAGGTGGAGCT

The qPCR amplification profile used an initial denaturation step of 95°C for 10m followed by 39 cycles of 95°C for 15s and annealing/extension at 60°C for 1m.

Works Cited

Alleman, Mary and Doctor, John. (2000) Genomic imprinting in plants: observations and evolutionary implications. *Plant Molecular Biology.* 43: 147-161.

Azevedo, R., Damerval, C., Landry, J., Lea, P.J., Bellato, C.M., Meinhardt, L.W., Guilloux, M.L., Delhaye, S., Toro, A.A., Gaziola, S.A. and Berdejo, B.D.A. (2003) Regulation of maize lysine metabolism and endosperm protein synthesis by opaque and floury mutations. *European Journal of Biochemistry.* 270 (24): 4898-4908.

Bagga, S., Adams, H.P., Rodriguez, F.D., Kemp, J.D., Sengupta-Gopalan, C. (1997) Coexpression of the maize delta-zein and beta-zein genes results in stable accumulation of delta-zein in endoplasmic reticulum-derived protein bodies formed by beta-zein. *Plant Cell.* 9: 1683-1696.

Below, F.E., Seebauer, J.R, Uribelarrea, M., Schneerman, M.C. and Moose, S.P. (2004) Physiological changes accompanying long-term selection for grain protein in maize. *Plant Breeding Reviews.* 24 (1): 133-151.

Bird, Adrian and Wolffe, Alan. (1999) Methylation-induced repression- Minireview Belts, Braces, and Chromatin. *Cell.* 99: 451-454.

Burr, B. and Burr, F. (1976) Zein synthesis in maize endosperm by polyribosomes attached to protein bodies. *PNAS*. 73: 515-519.

Campbell, R., Tour, O., Palmer, A., Steinbach, P., Baird, G., Zacharias, D., and Tsien, R. (2002) A monomeric red fluorescent protein. *PNAS.* 99 (12): 7877-7882.

Chaudhuri, Sumita and Messing, Joachim. (1994) Allele-specific parental imprinting of *dzr1*, a posttranscriptional regulator of zein accumulation. *PNAS.* 91: 4867-4871.

Coleman, C., Lopes, M., Gilliking, J., Boston, R. and Larkins, B. (1995) A defective signal peptide in the maize high-lysine mutant floury 2. *PNAS.* 92 (15): 6828-6831.

Coleman, C., and Larkins, B. The prolamins of maize. Eds. P. Shewry and R. Casey. Dordrecht: Kluwer Academic Publishers, 1999. 109-139.

Dudley, J.W., Clark, D., Rocheford, T.R. and LeDeaux, J.R. (2007) Genetic analysis of corn kernel chemical composition in the random-mated 7 generation of the cross of generations 70 of IHP x ILP. *Crop Sci.* 47: 45-57.

Duvick, D. (1961) Protein granules of maize endosperm cells. *Cereal Chemistry*. 38: 374-385.

Esen, A. (1986) Separation of the alcohol-soluble proteins (zeins) from maize into three fractions by differential solubility. *Plant Physiology.* 80: 623-627.

Feng, L.N., Zhu, J., Wang, G., Tang, Y.P., Chen, H.J., Jin, W., Wang, F., Mei, B., Xu, Zhengkai, Xu and Song, R. (2009) Expressional profiling study revealed unique expressional patterns and dramatic expressional divergence of maize alpha-zein super gene family. *Plant Molecular Biology.* 69: 649-659.

Fontes, E.B.P., Shank, B.B., Wrobel, R.L., Moose, S.P., Obrian, G.R., Wurtzel, E.T. and Boston, R.S. (1991) Characterization of an Immunoglobulin Binding-Protein Homolog in the Maize Floury-2 Endosperm Mutant. *Plant Cell.* 3: 483-496.

Forde, B.G., Heyworth, A., Pywell, J., Kreis, M. (1985) Nucleotide-Sequence of a B1-Hordein Gene and the Identification of Possible Upstream Regulatory Elements in Endosperm Storage Protein Genes from Barley, Wheat and Maize. *Nucleic Acids Research.* 13: 7327-7339.

Gehring, M., Choi, Y., and Fischer, R. (2004) Imprinting and seed development. *The Plant Cell.* 16: S203-S213.

Godwin, I. and Izquierdo, L. (2005) Molecular Characterization of a Novel Methionine-Rich δ-Kafirin Seed Storage Protein Gene in Sorghum (*Sorghum bicolor* L.). *Cereal Chemistry.* 82 (6): 706-710.

Heidecker, G., Chaudhuri, S., Messing, J. (1991) Highly clustered zein gene sequences reveal evolutionary history of the multigene family. *Genomics.* 10 (3): 719-732.

Holding, D.R., Otegui, M.S., Li, B.L., Meeley, R.B., Dam, T., Hunter, B.G., Jung, R. and Larkins, B.A. (2007) The maize floury1 gene encodes a novel endoplasmic reticulum protein involved in zein protein body formation. *The Plant Cell*. 19: 2569-2582.

Huang, S., Adams, W.R., Zhou, Q., Malloy, K.P., Voyles, D.A., Anthony, J., Kriz, A.L. and Luethy, M.H. (2004) Improving nutritional quality of maize proteins by expressing sense and anti-sense zein genes. *J. Agric. Food Chem*. 52: 1958-1964.

Hunter, B., Beatty, M., Singletary, G., Hamaker, B., Dilkes, B., Larkins, B., Jung, R. (2002) Maize opaque endosperm mutations create extensive changes in patterns of gene expression. *The Plant Cell.* 14: 1-22.

Hwang, Y., Ciceri, P., Parsons, R., Moose, S., Schmidt, R. and Huang, N. (2004) The Maize O2 and PBF Proteins Act Additively to Promote Transcription from Storage Protein Gene Promoters in Rice Endosperm Cells. *Plant Cell Physiology*. 45 (10): 1509-1518.

Kim, C., Woo, Y., Clore, A., Bunett, R., Carneiro, N. and Larkins, B. (2002) Zein Protein Interactions, Rather Than the Asymmetric Distribution of Zein mRNAs on Endoplasmic Reticulum Membranes, Influence Protein Body Formation in Maize Endosperm. *The Plant Cell.* 14: 655-672.

Langridge, P., Pintortoro, J.A., Feix, G. (1982) Zein Precursor Messenger-Rnas from Maize Endosperm. *Molecular & General Genetics.* 187: 432-438.

Larkins, B. and Hurkman, W. (1978) Synthesis and deposition of zein in protein bodies of maize endosperm. *Plant Physiology.* 62: 256-263.

Lauria, M., Rupe, M., Kranz, E., Pirona, Raul, Viotti, A., and Lund, G. (2004) Extensive maternal DNA hypomethylation in the endosperm of *Zea mays. The Plant Cell.* 16: 510-522.

Lin, B.-Y. (1984) Ploidy barrier to endosperm development in maize. *Genetics.* 107: 103-115.

Lund, G., Ciceri, P. and Viotti, A. (1995) Maternal-specific demethylation and expression of specific alleles of zein genes in the endosperm of *Zea mays*. *Plant J*. 8: 571-581.

Marks, D., Lindell, J., Larkins, B. (1995) Quantitative analysis of the accumulation of zein mRNA during maize endosperm development. *Journal of Biological Chemistry.* 260 (30): 16445-16450.

Marks, D., Lindell, J., Larkins, B. (1995) Nucleotide sequence analysis of zein mRNAs from maize endosperm. *Journal of Biological Chemistry.* 260 (30): 16451-16459.

Mohanty, A., Luo, A., DeBlasio, S., Ling, X., Yang, Y. Tuthill, D.E., Williams, K.E., Hill, D., Zadrozny, T., Chan, A., Sylvester, A.W., and Jackson, D. (2009) Advancing cell biology and functional genomics in maize using fluorescent protein-tagged lines. *Plant Physiology*. 149: 601-605.

Reggiani, R., and Soave, C. (1985) Factors affecting starch and protein content in developing endosperms of high and low protein strains of maize. *Genetica Agrarica.* 39: 221-232.

Russell, D.A. and Fromm, M.E. (1997) Tissue-specific expression in transgenic maize of four endosperm promoters from maize and rice. *Transgenic Res*. 6: 157-168.

Salas, A. (2008) Regulation of zein transgene expression in response to long term selection for grain protein concentration in maize. M.S. thesis. University of Illinois at Urbana-Champaign.

Schmidt, R., Ketudat, M., Aukerman, M., Hoschek, G. (1992) Opauque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. *The Plant Cell.* 4: 689-700.

Showalter, M.F. and Carr, R.H. (1922) Characteristic proteins in high- and low-protein corn. *J. Am. Chem. Soc*. 44: 2019-2023.

Song, R.T., Llaca, V., Linton, E. and Messing, J. (2001) Sequence, regulation, and evolution of the maize 22-kD alpha zein in gene family. *Genome Research.* 11: 1817-1825.

Song, R.T. and Messing, J. (2002) Contiguous genomic DNA sequence comprising the 19-kD zein gene family from maize. *Plant Physiology.* 130: 1626-1635.

Song, R.T., and Messing, J. (2003) Gene expression of a gene family in maize based on noncollinear haplotypes. *Proceedings of the National Academy of Sciences of the United States of America.* 100: 9055-9060.

Tsai, C.L. Genetics of storage protein in maize. Ed. J. Janick. *Plant Breeding Reviews: Volume 6.* AV Publication, 1989: 103-138.

Tsai, C.L., Dweikat, I. and Tsai, C.Y. (1990) Effects of source supply and sink demand on the carbon and nitrogen ratio in maize kernels. *Maydica* 35: 391-397.

Uribelarrea M, Below FE, Moose SP (2004) Grain composition and productivity of maize hybrids derived from the Illinois protein strains in response to variable nitrogen supply. *Crop Science*. 44: 1593-1600.

Uribelarrea, M., Moose, S.P., and Below, F.E. (2007) Divergent selection for grain protein affects nitrogen use efficiency in maize hybrids. *Field Crops Res*. 100: 82-90.

Vicente Carbajosa, J., Moose, S.P., Parsons, R.L. and Schmidt, R.J. (1997) A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences of the United States of America*. 94: 7685-7690.

Viotti, A., Sala, E., Marotta, R., Alberi, P., Balducci, C. and Soave, C. (1979) Genes and mRNAs coding for zein polypeptides in *Zea mays*. *Eur. J. Biochem.* 102: 211-222.

Wrage, E. (2005) Characterization of the alpha-zein storage proteins of the Illinois maize long-term selection strains. M.S. thesis. University of Illinois at Urbana-Champaign.

Woo, Y.M., Hu, D.W.N., Larkins, B.A. and Jung, R. (2001) Genomics analysis of genes expressed in maize endosperm identifies novel seed proteins and clarifies patterns of zein gene expression. *Plant Cell.* 13: 2297-2317.

Wyss, C., Czyzewicz, J. and Below, F.E. (1991) Source-sink control of grain composition in maize strains divergently selected for protein concentrations. *Crop Science.* 31 (3): 761-766.